Long-term plasticity of speech-related auditory functions in professional musicians

Thesis

presented to the Faculty of Arts

of

the University of Zurich

for the degree of Doctor of Philosophy

by

Cyrill Guy Martin Ott

of Zurich, Switzerland

Zurich, 2013

Accepted in the spring semester 2013

on the recommendation of

Prof. Dr. Lutz Jäncke

Prof. Dr. Volker Dellwo

ACKNOWLEDGEMENTS

In the following section i want to gratefully acknowledge all the people that contributed to this work and supported me during the entire Ph. D. period. First and foremost i want to thank my supervisor and mentor, Prof. Dr. Lutz Jäncke, as well as his wife, Dipl. Psych. Petra Jäncke, for inspiring and supporting me since the beginning of my academic career at the division of Neuropsychology at the University of Zurich. They offered me the opportunity to learn about different approaches of brain research and practical application of neuroscience, including techniques of neuroimaging, development and operationalisation of experimental paradigms as well as diagnostics in the realm of neuro- and traffic psychology. Moreover, i would like to particularly thank Prof. Dr. Lutz Jäncke for his supportive scientific supervision, inspiring discussions and constructive criticism of my work.

I also wish to express my thanks to all co-workers at the department of Neuropsychology at the University of Zurich for their instrumental and emotional support. In particular, many thanks go out to the doctors Kai Lutz, Jürgen Hänggi, Stefan Elmer, Mathias Oechslin, Pascale Sandmann, Michael Meier, Mike Brügger, Monika Kast, Yolanda Schlumpf, Andreas Pedroni, Ladina Bezzola, Stefanie Rotzer, Volker Baur and Franz Liem as well as the Ph. D. students Martina Hurschler, Jürg Kühnis, Katharina Rufener, Carina Klein and Lars Rogenmoser. Furthermore, i want to peculiarly thank Prof. Dr. Martin Meyer for accompanying and guiding me during the critical first phase and experiment in the context of this doctoral thesis. Similarly, i am specifically grateful to Dr. Nicolas Langer for letting me partake in his well-founded knowledge about electrical neuroimaging and MATLAB programming, thereby substantially contributing to my progress with respect to scientific research.

Another person to whom i would like to express my peculiar appreciation and deepest respect is my co-worker and Ph. D. student Gianclaudio Casutt. I would like to especially thank him for being so inspiring, supportive and patient.

Finally, my deepest thanks go out to my familiy and friends. I am notably indebted to my mother and my beloved wife for their ongoing support and encouragement as well as for giving me so much love and emotional backup during my Ph. D. period. And last but not least my warmest thank goes to my little daughter for filling my life with sunshine and joy.

Original research articles included in this doctoral thesis

Experiment 1:

Ott, C. G., Langer, N., Oechslin, M. S., Meyer, M., and Jäncke, L. (2011). Processing of Voiced and Unvoiced Acoustic Stimuli in Musicians. *Front Psychol* 2 195.

Experiment 2:

Ott, C. G., and Jäncke, L. (2013). Processing of self-initiated speech sounds is different in musicians. (accepted for publication in *Front Psychol*).

Experiment 3:

Ott, C. G., Stier, Ch., Herrmann, Ch., and Jäncke, L. (2013). Musical expertise affects attention as reflected by auditory evoked gamma-band activity in human EEG. (submitted to *Neuroreport*).

Table of contents

I. Summary

An abundance of previous studies have shown that musical expertise and intense musical training are associated with rather widespread short- and long-term functional and structural alterations in the architecture of the human brain (for a review, see e.g. Herholz and Zatorre, 2012; Jäncke, 2009b; Peretz and Zatorre, 2005; Shahin, 2011). In the course of these discoveries, the question whether such neuroplastic effects related to musicianship can also be found with respect to the processing of speech has been increasingly addressed in recent years. Most of the corresponding research indeed found convergent evidence for altered auditory functioning due to musical training to transfer into the domain of speech (e.g. Elmer et al., 2012; Fedorenko et al., 2009; François et al., 2012; Magne et al., 2003; Magne et al., 2006; Parbery-Clark et al., 2011b; Patel, 2003; Rogalsky et al., 2011; Sammler et al., 2009; Strait et al., 2012; Wong et al., 2007). Insofar musicians can be considered not only an excellent model to study conventional learning-induced brain plasticity, but also to be predestinated to investigate cross-domain transfer effects associated with intense long-term training. However, most of these studies focused on supra-segmental, spectral and higher-order aspects of speech such as the processing of pitch, timbre, prosody and syntax.

Based on these pre-existing findings regarding cross-domain neuroplasticity in professional musicians, the present doctoral thesis aimed to further elucidate long-term reorganization in the realm of speech-related low-level auditory functions associated with musical expertise. Thereby, possible alterations with respect to the processing of rapidly changing spectro-temporal information as well as sensory-motor interplay and sound intensity in the speech signal / auditory stream as a function of musicianship were of particular interest. Consequently, I planned and performed a series of EEG studies specifically designed to address these issues.

In **experiment 1**, we found that professional musicians process unvoiced acoustic stimuli differently than non-musicians. In particular, musically experienced subjects seem to use the same network to analyze unvoiced stimuli as they use for voiced ones, whereas different networks seem to be recruited in musical laymen. Moreover, musicians seem to devote more neurophysiological resources into the analysis of unvoiced segments.

The results of **experiment 2** indicate that so-called internal forward models (e.g.

Wolpert and Miall, 1996) not only suppress cortical responses to incoming stimuli, but also facilitate processing of sensory (sound) events in terms of faster processing times as reflected by shorter microstate durations during early stages of auditory analysis. Moreover, we could demonstrate that musicians seem to profit from such facilitation also when analyzing speech, while non-musicians did not show such a cross-domain facilitation effect.

Eventually, we could show in **experiment 3** that intense musical training also exerts influence on top-down processes as reflected by differential effects of attention between musicians and non-musicians on the intensity-modulated early auditory evoked gamma-band response (GBR) with respect to inter-trial phase-coherence.

Taken together, the results reported in this doctoral thesis provide further evidence that music and speech share neuronal resources. Moreover, they corroborate and complement previous studies showing that functional alterations associated with musicianship do not only manifest in exceptional acuitiy in music processing, but also transfer to speech-related auditory functions.

II. Zusammenfassung

Eine Vielzahl wissenschaftlicher Studien belegen, dass musikalische Expertise und intensives musikalisches Training mit weitreichenden kurz- und langfristigen strukturellen sowie funktionellen Veränderungen in der Architektur des menschlichen Gehirns verbunden sind (für einen entsprechenden Überblick siehe z.B. Herholz and Zatorre, 2012; Jäncke, 2009b; Peretz and Zatorre, 2005; Shahin, 2011). Im Zuge dieser Entdeckungen wurde in den letzten Jahren vermehrt der Frage nachgegangen, inwiefern sich solche neuroplastischen Vorgänge in Zusammenhang mit musikalischem Training auch auf die Verarbeitung von Sprache auswirken. Tatsächlich verdeutlichen die Ergebnisse aus der entsprechenden Forschung, dass sich auch die Funktionsweise der auditorischen Verarbeitung von sprachlichen Signalen als Folge intensiven musikalischen Trainings verändert (z.B. Elmer et al., 2012; Fedorenko et al., 2009; François et al., 2012; Magne et al., 2003; Magne et al., 2006; Parbery-Clark et al., 2011b; Patel, 2003; Rogalsky et al., 2011; Sammler et al., 2009; Strait et al., 2012; Wong et al., 2007). Insofern können professionelle Musiker nicht nur als vorzügliches Modell betrachtet werden, um die domänenspezifische Plastizität des menschlichen Gehirns in Zusammenhang mit intensivem Training zu studieren, sondern auch, um domänenübergreifende kortikale Reorganisationsprozesse genauer zu erforschen. Allerdings ist an dieser Stelle zu erwähnen, dass ein Grossteil der bisherigen Studien zu den Auswirkungen musikalischen Trainings auf sprachliche Funktionen ihren Fokus auf supra-segmentale, spektrale sowie höhere Ebenen der Sprachverarbeitung gelegt haben (bspw. Tonhöhe, Klangfarbe, Prosodie und Syntax).

Basierend auf diesen Ergebnissen die domänenübergreifende Neuroplasitizität bei professionellen Musikern betreffend, zielt die vorliegende Doktorarbeit darauf ab, das vorhandene Wissen über die Auswirkungen intensiven musikalischen Trainings auf die elementare, sprachbezogene auditorische Verarbeitung weiter zu vergrössern. Dabei liegt der Fokus insbesondere auf den konkreten Manifestationen domänenübergreifender Reorganisationsprozesse bezüglich der Prozessierung von sich schnell ändernder, spektrotemporaler Information, der Intensität auditorischer Signale sowie der sensumotorischen Interaktion. Dementsprechend führte ich im Rahmen dieser Doktorarbeit eine Reihe wissenschaftlicher Experimente durch, welche gezielt im Hinblick auf diese Fragestellungen entwickelt wurden.

Die Ergebnisse von **Experiment 1** haben gezeigt, dass professionelle Musiker stimmlose akustische Stimuli anders als Laien verarbeiten. Dabei scheinen Musiker das gleiche Netzwerk für die Analyse stimmloser Stimuli zu rekrutieren, wie sie dies für stimmhafte Stimuli tun. Im Gegensatz dazu scheinen Laien für die entsprechenden Analysen zwei unterschiedliche Netzwerke zu verwenden. Eine weitere Möglichkeit zur Erklärung der entsprechenden Ergebnisse besteht darin, dass Musiker mehr neurophysiologische Ressourcen in die Analyse stimmloser Stimuli investieren, um diese möglichst genau zu analysieren.

In **Experiment 2** konnten wir nachweisen, dass sogenannte „internal-forward Modelle" (siehe z.B. Wolpert and Miall, 1996) nicht nur die kortikale Antwort auf eingehende, selbst-initiierte akustische Reize abschwächen, sondern gleichzeitig auch deren Verarbeitung, im Sinne kürzerer Prozessierungszeiten während frühen Verarbeitungsstadien, erleichtern. Des Weiteren konnten wir zeigen, dass Musiker auch in Bezug auf die Verarbeitung selbst-initiierter sprachlicher Signale von diesem Erleichterungseffekt profitierten, während bei Laien kein solch domänenübergreifender Erleichterungseffekt zu beobachten war.

Die Ergebnisse von **Experiment 3** wiederum ergaben deutliche Hinweise darauf, dass intensives musikalisches Training auch einen Einfluss auf top-down Prozesse ausübt. Dies war anhand unterschiedlicher Effekte von Aufmerksamkeit in Abhängigkeit des musikalischen Expertenstatus auf die inter-trial Phasenkohärenz der durch akustische Stimuli verschiedener Intensitäten evozierten gamma-band Aktivität zu erkennen.

Zusammenfassend lässt sich demnach festhalten, dass die hier vorgestellten Ergebnisse der im Rahmen dieser Doktorarbeit durchgeführten Studien weitere Belege dafür liefern, dass Musik und Sprache auf gemeinsame neuronale Ressourcen zurückgreifen. Zudem bestätigen und ergänzen sie die bisherige Forschung in Zusammenhang mit musikalischem Training und domänenübergreifenden neuroplastischen Reorganisationsprozessen, indem sie verdeutlichen, dass sich funktionelle Veränderungen von auditorischen Verarbeitungsprozessen als Folge musikalischer Expertise auch auf verschiedenen Ebenen sprachbezogener auditorischer Funktionen manifestieren.

1. Introduction

1.1. Brain plasticity and neuronal reorganisation

For a large period of time the consensus among researchers was that cortical and subcortical brain structures are static and fairly unalterable after a „critical period“ of strong plasticity whithin a specific age range. At the same time, brain regions related to memory functions such as hippocampal areas were believed to be more susceptible to functional and structural reorganization processes also beyond this pivotal phase. This concept of the „critical period“, where a certain fixation of neural pathways and structures takes place, was widely accepted (Jäncke, 2009a). Though, a vast amount of studies conducted with various methodologies changed this understanding substantially over the past two decades, showing that almost the entire human brain exhibits structural and functional reorganisation throughout the entire lifespan in response to experience (for a review, consider e.g. Bavelier et al., 2012; Buschkuehl et al., 2012; Gao et al., 2012; Jäncke, 2009a). As a consequence, different organization principles of neuroplasticity on a molecular, cellular, physiological as well as on a micro- and macroanatomical level are known to date.

In the following section, i will briefly describe a frequently investigated and well-known principle of neuronal plasticity as a representative for neuroplastic processes. Afterwards, the beginnings of research on brain plasticity in humans are shortly introduced and different investigation techniques are enumerated.

1.1.1. Hebbian learning as an exemplary principle of brain plasticity

In 1949, canadian neuropsychologist D. O. Hebb proposed the „Hebbian theory of learning“ (Hebb, 1949), an actual breakthrough in regard to the concept of neuroplasticity. In particular, this theory assumes that synapses strengthen with use and that simultaneously firing neurons also preferentially interconnect. To exemplify this basic principle, D. O. Hebb devised the following statement about neuroplasticity’s mode of operation: „When an axon of cell A is near enough to excite cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A’s efficiency as one of the cells firing B, is increased“ (Hebb, 1949).

With this so-called „Hebb’s law“ as a starting point, knowlegde about mechanisms

and principles of brain plasticity rapidly increased. For example, the discovery of long-term potentiation (LTP) several years later amended Hebb's theory and marked another milestone in the understanding of neuronal reorganization processes (e.g. Lømo, 1966; Bliss and Lomo, 1973; Douglas and Goddard, 1975; Lømo, 2003). In particular, LTP constitutes one of several phenomena underlying the modulation of synaptic strength and is characterized by a long-lasting enhancement in signal transmission between two neurons that results from their synchronous stimulation (Cooke and Bliss, 2006). This form of a cellular learning mechanism was first discovered in the connections from the perforant pathway to the dentate gyrus of rabbits by Terje Lømo in 1966. When stimulating pre-synaptic fibers of the perforant pathway with high-frequency electrical pulse-trains, Lømo observed that the postsynaptic cells' response to subsequent single-pulse stimuli could be enhanced for a long period of time (Lømo, 1966). Interestingly, application of electrical low-frequency pulse-trains to the pre-synaptic fibers induces so-called long-term depression (LTD), which can be understood as the opposing process to LTP. It serves to selectively weaken specific synapses in order to make constructive use of synaptic strengthening caused by LTP. This becomes necessary as synapses would ultimately reach a ceiling level of efficiency when being allowed to unrestrictedly continue increasing in strength, which in turn would inhibit the encoding of new information (Purves et al., 2008). The molecular mechanisms of LTP and LTD have been shown to depend on the properties of the doubly gated N-methyl-D-aspartate (NMDA) receptor located on the dendritic spines of postsynaptic neurons that show LTP/LTD and its agonist glutamate. In addition, the functional properties of calcium and magnesium ions also appear to be critical in forming LTP/LTD (Gazzaniga et al., 2002).

Beside modulating the strength of synaptic transmission, LTP and LTD can also shape the physical connections between neurons. Driven by mRNA / DNA cascades of molecular processes following calcium influx on the dendritic receptors, dendritic spines appear and disappear in the latency range of seconds during LTP/LTD. As a consequence, the surface area which is available on the dendrites for building contacts with axon terminals is either increased or reduced (e.g. Tashiro and Yuste, 2003).

1.1.2. Studying neuroplastic changes in the human brain

As most of the early studies between the late 1970s and early 1980s investigating cortical plasticity examined neuronal reorganisation in animals, Ramachandran's work on patients suffering from phantom limb pain after amputation of hands, arms or legs

can be considered as a trailblazer for the study of brain plasticity in humans (Jäncke, 2009a). Based on Ramachandran's research, phantom limb pain is thought to result from disorganization in the somatosensory cortex and the loss of input from the targeted area. This assumption receives strong support by the phenomenon that phantom limb subsides after physical remapping and adaptation of the somatosensory representation to the new situation has been completed. In particular, Ramachandran and his coworkers showed in a series of studies that the somatotopic representation of the body parts that are represented nearby the dismembered extremity reintegrate the cortical structures that formerly coded the sensory perception of the amputated limb (Ramachandran et al., 1992; Ramachandran et al., 1995; Ramachandran and Rogers-Ramachandran, 2000).

Since this early work, research and knowledge on brain plasticity in humans has been constantly increasing, not least due to the emergence and further development of imaging techniques like magnetic resonance imaging (MRI), electro- and magnetoencephalography (EEG/MEG). These imaging techniques provide various application possibilities such as measuring electrical brain activitiy directly on the scalp in form of voltage fluctuations (EEG), correlative measurements of brain activity (blood oxygen saturation, fMRI-BOLD), morphometric analyses (gray and white matter density, volume, cortical thickness and surface area) or the identification and mensuration of white matter fiber tracts (diffusion tensor imaging, „DTI"). In combination, fMRI, MRI, DTI and EEG/MEG constitute very powerful tools to uncover functional and structural neuroplastic changes in the living human brain. As a consequence, training and experience induced alterations in the architecture and functionality of neuronal assemblies have not only been demonstrated in healthy adults (e.g. Buschkuehl et al., 2012; Draganski and May, 2008; Jäncke, 2009a; Wan and Schlaug, 2010), but also in children (e.g. Hyde et al., 2009; Meyer et al., 2011; Schlaug et al., 2009; Trainor et al., 2003) and elderly people (e.g. Engvig et al., 2010; Lövdén et al., 2010; Parbery-Clark et al., 2011a).

In conclusion, today there is a wide consensus among researchers that the brain exhibits the ability to change itself as a function of exposure to a particular environment, experience or training. These reorganization processes and their underlying principles are the subject of ongoing investigations on molecular, cellular, physiological as well as micro- and macro-anatomical levels by using different methods and techniques.

1.2. Professional musicians as a model of neuroplasticity

Playing an instrument involves the interplay of different sensory systems and the motor system while at the same time making high demands on a variety of higher-order cognitive functions such as working memory and attention. By providing such complex and challenging preconditions, musical training offers an excellent opportunity to investigate how sensory-motor systems interface with cognition and how different types of training influence these interactions. In this regard, musical training emerges to be especially advantageous and fruitful. As playing music requires fine-grained perception and precise motor control that is unlike any other everyday activity, confounding influences of other types of experience are substantially reduced. In addition, professional musicians dedicate enormous amounts of time to their instrumental practice, while at the same time it is relatively easy and safe to administer auditory and/or motor training in a musical context in a laboratory or clinical environment. Hence, musical training also allows to study both short- and long-term neuronal plasticity. Finally, linking changes in brain function and structures to altered behavior is readily accessible, as the behavioral consequences of musical training can be directly measured by using either psychophysics and/or cognitive tasks.

In the following sections i will explicate some representative research investigating brain plasticity in the auditory system and the motor network induced by musical training. Subsequently, exemplary transfer effects of musical training on speech-related functions will be described.

1.2.1. Musical training and the auditory system

Due to its upmost and inherent importance for music, one of the systems that is most altered by musical experience is the auditory system. Functional and structural reorganization in consequence of exposure to musical training already takes place at very early stages in the auditory pathway and extends up to auditory-related neocortical association areas. Corresponding effects have been found in the brainstem (e.g. Bidelman et al., 2011b; Bidelman et al., 2011a; Wong et al., 2007), in primary and surrounding auditory-related cortex (e.g. Bermudez et al., 2009; Schneider et al., 2002; Schneider et al., 2005) as well as in areas involved in higher-order auditory cognition (e.g. Lappe et al., 2008; Sluming et al., 2002).

By using music as both an active training protocol as well as a stimulus in the

context of mere auditory training, researchers can exploit the full potential of musical training at investigating brain plasticity. Doing so, it is possible to differentiate the effects of active vs. passive training and to scrutinize the roles and interactions of the involved cortical networks.

First, it is important to point out some general features of well-known changes of auditory cortical responses due to short- and long-term exposure to specific, structured sounds. To begin with, it is known mainly from animal studies that there are long-lasting changes to receptive field properties, temporal aspects and an expansion/reduction of tonotopically organized cortex as a function of exposure to appropriate stimuli (e.g. Ahissar et al., 1998; Bao et al., 2004; Polley et al., 2006; Shetake et al., 2012). In addition, these changes appear to depend on the particular task (e.g. Ohl and Scheich, 2005; Polley et al., 2006), and are strongest when auditory input is behaviorally relevant and tasks are actively trained (e.g. Fritz et al., 2005; Ohl and Scheich, 2005; Recanzone et al., 1993). Furthermore, reorganization of cortical mapping and neuronal tuning strongly depends on the reward value of a learned stimulus (e.g. David et al., 2012) and the maturational state of the nervous system (e.g. de Villers-Sidani et al., 2008). Last but not least, there are also dynamic and reversible short-term alterations in neural response properties reflecting specific task contingencies (e.g. Fritz et al., 2005).

Even without any special musical education, humans clearly have a good sense of music and show sensitivity to musical relationships like tonality and rhythm (e.g. Hannon et al., 2004; Toiviainen and Krumhansl, 2003). Research in musically untrained subjects correspondingly suggests that passive exposure to music alone can be sufficient to alter the neuronal response to musical sounds to some extent (e.g. Koelsch et al., 2000; Wong et al., 2009). However, the most pronounced alterations in functional and structural architecture of the auditory system can be seen in conjunction with extensive musical training that involves active playing of an instrument, a prerequisite that is perfectly met by professional musicians.

A good example for such prominent neuroplastic changes in the auditory system of highly trained musicians comes from a study of Schneider and colleagues (Schneider et al., 2002). Using magnetoencephalograhy, these authors showed that the neurophsysiological activity in the primary auditory cortex was more than 100% larger 19-30ms after tone onset in musicians compared to non-musicians. Furthermore, grey matter volume of the anteromedial part of Heschl's gyrus was 130% larger in musicians. Interestingly, both quantities were highly correlated with musical aptitude, indicating

that both the morphology and neurophysiology of Heschl's gyrus have a substantial impact on musical skills. The same group performed another study with results not less astonishing (Schneider et al., 2005). Here, Schneider et al. (2005) found a strong relationship between anatomical features in the primary auditory cortex of professional musicians and the strategy they used to process complex tones. In particular, musicians preferentially analysing the fundamental pitch (f0 or F0, i.e. the lowest frequency in a harmonic spectrum) of complex tones showed a leftward asymmetry of grey matter volume in Heschl's gyrus, whereas the corresponding grey matter volume asymmetry of those preferring to analyse the spectral pitch was shifted rightwards. Again, an anatomical feature of the auditory system correlated with a particular behavioral peculiarity within a group of musically proficient experts.

But marked differences between musically trained and untrained subjects can not only be observed with respect to anatomical measurements, as shown by a myriad of studies rather emphasizing functional differences in the auditory system. For instance, (Lappe et al., 2008; Lappe et al., 2011) found cortical auditory representations of musical tones to be enhanced in subjects that practiced piano playing for two weeks, as indicated by enlarged musically elicited mismatch negativities from MEG measurements. Other research groups have demonstrated e.g. increased gamma-band responses reflecting improved binding of auditory features (Shahin et al., 2008), increases in auditory responses to unexpected tones (Fujioka et al., 2004) and more efficient neural detection of pure and harmonic tones as suggested by longer EEG P1 latencies to pure and decreased P1 amplitudes to harmonic tones as a function of musical training.

1.2.2. Musical training and the motor network

In 2006, Bangert and Schlaug found that pianists show a larger hand motor area on both hemispheres (the so-called „omega-sign"), while violinists typically showed the omega-sign only on the right hemisphere that controls the left hand (Bangert and Schlaug, 2006). A similar finding was reported three years earlier by (Lotze et al., 2003). Here, professional musicians showed enhanced electromyographic (EMG) amplitudes and more focused cerebral activation during movement execution and imagery compared to non-musicians. In contrast, representations of the vocal tract are enhanced in singers (Kleber et al., 2007). These specific anatomical features clearly demonstrate that active musical performance relies on a distributed motor network which exhibits specifity to

the type of action involved in playing a certain instrument. Other parts of the motor network that differ anatomically between musicians and musical laymen include the anterior corpus callosum (e.g. Schlaug et al., 1995), motor and premotor cortex (e.g. Bermudez et al., 2009; Gaser and Schlaug, 2003a; Gaser and Schlaug, 2003b) as well as the cerebellum (e.g. Hutchinson et al., 2003).

There is also ample evidence for musical training induced functional differences in the sensorimotor system. Primary and secondary motor areas (M1, SMA, pre-SMA and CMA) have been shown to be considerably less activated during complex bimanual movements in professional pianists than in non-musicians (Jäncke et al., 2000). A similiar finding has been reported by (Koeneke et al., 2004). These authors observed reduced cerebellar hemodynamic responses in professional pianists compared to musical laymen during complex uni- and bimanual finger movements. Taken together, these findings suggest that due to long-term motor practice the involved motor areas become more efficient and fewer neurons need to be recruited to perform an accordant task. On the other hand, (Bangert et al., 2006) found increased neuronal activity in pianists and a distinct predominantly left-hemispheric musicianship-specific network to be co-activated during either passive listening to short piano melodies or pressing keys on a mute piano keyboard. The concerned networks included dorsolateral and inferior frontal cortex, superior temporal gyrus, supramarginal gyrus and supplementary motor and premotor areas. Considering the different activiation patterns in premotor and motor areas (less vs. more activation) in comparison to the aforementioned studies (Jäncke et al., 2000; Koeneke et al., 2004), it seems that the motion-related demands as well as the musical relevance of a certain task plays an important role on how cortical resources in the motor system are recruited in musical experts.

1.2.3. Transfer effects from musical training on elementary speech-related functions

Expressive and receptive musical functions have repeatedly been shown to recruit a distributed perisylvian network that at least partially overlaps with regions involved in elementary speech processing (e.g. Brown et al., 2006; Rogalsky et al., 2011; Sammler et al., 2009; Wilson et al., 2011). As a consequence, the question arises whether and to what extent musical training transfers to basic processing of speech.

In order to transfer enhanced auditory skills to basic speech functions, musical training has to target and modify the particular neuronal assemblies that also underlie

elementary speech processing in such a way that their functioning is optimized and improved. As music and speech processing share several mechanisms from a psychoacoustics point of view, chances are high that such modifications in fact take place. For instance, the auditory system has to encode pitch or voice cues as well as rhythm in order to extract meaning from ongoing speech or a piece of music. Moreover, segregation of a voice from background noise strongly resembles the separation and identification of a particular musical instrument within a symphonic orchestra or jazz band. Considering these exemplary similarities, musical training actually should also recruit resources associated with speech processing and thereby enhance the neuronal mechanisms underlying elementary speech functions and vice versa. Corresponding evidence for such mutual influence comes from a wide variety of studies using different methodologies and approaches.

Training-related changes in brain regions relevant for music and elementary speech processing have been shown already at the level of the brainstem (e.g. Krishnan et al., 2009), as evidenced by training-induced alterations in the EEG frequency following response (FFR). The FFR reflects the amplitude and phase-alignment of neuronal firing with a sound's fundamental frequency (F0) in the brain-stem. For instance, (Musacchia et al., 2007) found earlier and stronger FFR's in musicans to both speech and music stimuli, reflecting enhanced representation of F0 in the brain-stem. (Song et al., 2008) showed the FFR to be augmented after tonal training in English speakers for Mandarin tones, again indicating that specific training either in both the domain of speech or music can induce neuroplastic modifications in the brain-stem FFR. Moreover, these functional alterations were consistent with better pitch discrimination abilities (Musacchia et al., 2007; Musacchia et al., 2008; Wong et al., 2007) and more precise pitch encoding of speech sounds under passive listening conditions in musicians compared to non-musicians (Wong et al., 2007). Taken together, these results clearly show that inter-domain functional transfer already at processing levels as early as the brain-stem is fairly probable.

Whereas training in both speech and musical realms targets and demonstrably alters coinciding brain-stem functions, evidence for such an overlap at the level of primary auditory cortex (A1) is less clear. One prominent measurement indexing activity in A1 is the middle latency response (MLR) measured in EEG and MEG. MLR can be measured as early as 10 to 50 ms after sound onset and is believed to reflect thalamo-cortical input to and processing in A1 (Hall, 2006). Indeed, larger MLR components were

shown to denote improved pitch and rhythm encoding in musicians relative to non-musicians for pure tones and music sounds (e.g. Neuhaus and Knösche, 2008; Schneider et al., 2002; Shahin et al., 2004). On the other hand, discrimination training on voice onset time (VOT) of speech syllables in young adults resulted in smaller MLR's but enhancement of the later occurring N1 and P2 auditory evoked potential (AEP) components (Tremblay and Kraus, 2002). Moreover, (Hertrich et al., 2000) showed reduced MLR for periodic (containing pitch information) compared to aperiodic (lacking pitch information) speech-like stimuli, whereas the activation pattern of the later occuring M100 (the magnetic counterpart of the EEG N1) was the other way round. As N1, P2 and M100 have sources also in non-primary auditory cortex (NPAC; e.g. Atcherson et al., 2006; Shahin et al., 2003; Zouridakis et al., 1998), these results suggest that experience-based improvements in speech-processing may originate on a higher cortical level than that of music. Accordingly, musical expertise-related effects modifying elementary speech processing should emerge somewhat later and thus be predominantly observable in the N1 (~100 ms) and P2 (~180 ms) latency range of AEP's instead of the MLR.

Corroborating findings with respect to this assumption comes from a variety of studies, primarily showing comparable experience-dependent modulations of the N1 and P2 components as a function of musical or speech-related training. For instance, P2 enhancement can be observed after musical (Pantev et al., 1998; Shahin et al., 2003) as well as after vowel and VOT discrimination training (Reinke et al., 2003; Tremblay and Kraus, 2002). As these results clearly indicate that musical training and training in the speech domain impact a shared neural system at this level similarly, functional transfer to speech processing following musical training can be inferred. It is thus not surprising that the neuroplastic P2, wich has been related to temporal acuity (Tremblay et al., 2001), has also been associated with an improved ability of musicans to encode metric structure in speech as reflected in stronger P2 amplitudes coupled with better accuracy in a corresponding forced-choice discrimination task (Marie et al., 2011). Taken together, the results of these studies suggest that P2 enhancement due to musical experience reflects improved coding of the metric structure in speech and coincides with enhanced temporal acuity in speech processing. As a consequence, it seems legit that a transfer effect for improved processing of spectrotemporal features in speech can be facilitated by musical training at the level of NPAC (Shahin, 2011).

2. Methods

2.1. Electroencephalography (EEG)

The EEG technique is relatively easy to apply and enables direct measurement of electrical brain activity. By placing electrodes on the outer surface of the scalp, voltage fluctuations originating from synchronous activity of neuronal assemblies can readily be recorded. In contrast to other imaging techniques like fMRI, EEG provides an excellent temporal resolution on a time-scale of milliseconds. This particular feature makes it optimally suited to investigate the functional properties and temporal dynamics of the fast-paced neuronal processes associated with speech-related cortical activity. Since the original research in the present thesis focused on transfer effects of musical training on early speech-related auditory functions, the EEG technique emerged as the method of choice and was correspondingly used in all three studies. Moreover, EEG offers a wide variety of possible ways to analyze the obtained data, ranging from classical component analysis of event-related potentials at single electrodes to rather sophisticated evaluations of scalp map topographies, time-frequency decompositions or functional connectivity. In the following sections i will first give a short overview over the origins of the EEG signal and then briefly describe the analysis methods that are used in the original research included in this thesis. However, for a detailed description of the methods used, i refer to the respective method sections in the original publications as well as the corresponding technical literature (e.g. Hall, 2006; Michel et al., 2009; Murray et al., 2008).

2.1.1. Origins of the EEG signal

The EEG signal reflects the immediate mass action of neuronal populations organized in large-scale networks from a wide range of brain systems and provides a specifically direct and integrative noninvasive insight into human brain function. Due to its unique high temporal resolution, it does not only concern the possible neuronal generator of a specific scalp potential at one given moment in time, but also the possible generators of rhythmic oscillations in different frequency ranges. However, multichannel intracranial recordings have shown that spike amplitudes of regularly firing neurons fall of rapidly outside a radius of 50 micrometers from their origin, which limits their spatial spread to the sub-millimeter scale (Henze et al., 2000). Moreover, the short spike duration

minimizes the probability of summation over time. This prevents local field potentials due to individual spikes from becoming measurable in the EEG on the scalp, which is at least 2 cm away from actual brain tissue. Thus, individual spikes or action potentials that propagate along white matter structures can be neglected as direct EEG generators. The major generators of the scalp EEG are rather extended patches of gray matter organized in cortical columns, polarized through synchronous synaptic input either in an oscillatory fashion or as tranisent evoked activity (Michel et al., 2009). Such columns contain thousands of large pyramidal cells that are aligned perpendicularly to the cortical surface and exhibit extensive synaptic interconnections from different structures. These synaptic connections are involved in the transmission of chemical and/or electrical signals between two or more cells. A very simplified description of this transmission process and in succession the generation of the EEG signal is as follows: Action potentials travelling along the axon of a neuron induce the release of neurotransmitter into the synaptic cleft as they arrive at the axon terminal. After passing the synaptic cleft, these neurotransmitters connect to receptors in the post-synaptic cell membrane, thereby changing their molecular structure and influencing the flow of electrically charged ions through the post-synaptic membrane. As a consequence, electrical potentials are differentially distributed at the dendritic membrane in such a way that the intracellular space becomes more positive relative to the extracellular space. At the same time, the potential differences at the soma of the neuron shows exactly the opposite pattern. These differently orientated polarities can be taken as the origin of electromagnetic dipoles, which are orthogonally orientated to the surface of the cortex. Subsequent changes in strength and orientation of these dipoles can then be recorded by electrodes placed on the scalp surface (Elmer, 2010).

2.1.2. Event-related potentials (ERP)

The term „event-related potential“ or ERP summarizes all electrocortical potentials that can be measured in the EEG before, during and after any sensory, motor or cognitive event. The magnitude of ERP's is rather small (~1-40 μV), thus they are generally not detectable in raw EEG measurements. Based on the assumption that the neurophysiological cortical processes recorded in response to repeated sensory, motor or cognitive stimulation are more or less constant whereas background EEG activity is randomly distributed, ERP's are computed by averaging EEG segments time-locked (i.e. „event-related“) to repeated identical stimulation with sensory, motor or cognitive

events. The more segments are available for averaging, the better the ERP is emerging from EEG background noise. After this averaging procedure, the resulting ERP mostly exhibits a series of multiple distinct positive and negative deflections with characteristic spatiotemporal dynamics that reflect different neurophysiological and psychological processes. These deflections can then be analysed in terms of amplitude and latency variations as a function of different conditions. With respect to ERP's, the present work mainly focuses on the elements of a specific auditory event-related composite of deflections with middle to long latency (~100 – 200 ms), namely the prominent N1/N100 and P2/P200 components, also known as the N1-P2 complex. In particular, in experiments 1 & 2 we analysed the auditory N1 component, whereas in experiment 3 we analysed the auditory N1-P2 complex.

In general, the auditory N1 and P2 components encode low-level acoustical features such as sound onset and pitch and also represent higher-level sound features brought about by the spectrotemporal complexity of speech and music (Shahin, 2011). Moreover, notably the N1 component can be modulated by (selective) attention (Hillyard et al., 1973). Both the N1 and P2 components are believed to originate in A1 and its surrounding areas (Bosnyak et al., 2004; Shahin et al., 2003), including belt and parabelt regions of the superior temporal gyrus (Hackett et al., 2001). Due to the mainly vertical orientation of the cortical columns within these specific supratemporal regions, the auditory N1 and P2 components exhibit their maximal magnitude at central scalp sites and are thus preferrably obtained at the central electrode (Cz) located at the vertex of the head.

2.1.3. Microstate analysis (Topographic pattern analysis / EP-mapping)

Specific ERP components such as the N1 and P2 are typically obtained and analysed at a single electrode or pre-defined pooled subsets of EEG sensors at rather narrowly circumscribed scalp sites. Moreover, experimenters usually look for particular components within a-priori defined time windows and pick the peaks (and troughs) of these, thereby increasing the possibility that other time periods and effects are overlooked. Finally, the morphology of ERP waveforms has been shown to vary depending on the chosen reference for EEG recording and evaluation, representing another clear-cut source of experimenter bias (e.g. Murray et al., 2008). Hence, the information content provided by such an analysis is relatively limited and rather biased by various decisions to be made by the experimenter. In contrast, analysis methods

relying on scalp map topography are reference-free and exploit the entire information available in the EEG signal by accounting for the input from all applicable electrodes at any given time point. Moreover, they relieve the experimenter from picking peaks whithin pre-defined latency windows at constricted subsets of electrodes, thereby substantially decreasing experimenter introduced bias in the evaluation of EEG data (e. g. Michel et al., 2009; Murray et al., 2008).

Hence, in experiments 1 and 2 we used the so-called „topographic pattern analyis“ or „EP-mapping“ approach. This method derives from the principle of functional microstates, which was first introduced by Dietrich Lehmann (e. g. Lehmann, 1987). This concept is based on the observation in both continuous EEG and ERP's that EEG map series (electric potential field configurations) do not vary randomly as a function of time, but are rather characterized by sequential periods of stable map topographies with brief intervening intervals of topographic instability. Each of these functional microstates can be considered to represent a basic building block of information processing (Michel et al., 2009). Topographic pattern analysis / EP-mapping refers to the procedures used to identify the periods of topographic stability within and between experimental conditions and / or groups. The aim of these procedures is to obtain the particular sequences of stable map topographies (i.e. „microstates“) in varying conditions, which then can be statistically compared with respect to different parameters such as the duration and the timepoint of first on- or last offset of any given microstate.

The particular method that was used for EP-mapping in the original research of the present work is a modified (topographic) agglomerative hierarchical clustering procedure („AAHC“ or „T-AAHC“). It has been specifically designed for the analysis of EEG/ERP's and is described in detail in Murray et al. (2008) or Michel et al. (2009). Briefly summarized and simplified, AAHC / T-AAHC identifies the dominant mean map topographies (i.e. „template maps“) that best describe the group-averaged ERP's across all conditions and groups at any given time point. The clustering algorithm does not account for the latencies of these template maps but only for their topographies. This is done as a hypothesis generation step within which the sequence of template maps that best accounts for the data is determined. In a next step, these hypotheses generated at the group-averaged level are statistically tested by means of a fitting procedure that is based on the spatial correlation between the template maps obtained from the group-averaged ERP's and the single-subject ERP data. This so called „single-subject fitting“ procedure returns several different dependent measures such as map/microstate

duration, global explained variance (GEV), first-onset and last-offset (for an overview see Murray et al., 2008). These measurements can then be extracted and further analysed by means of any applicable statistical procedure.

2.1.4. Time-frequency analysis (wavelet analysis)

Besides specific single scalp potentials obtainable at one given moment in time, the EEG signal also contains rhythmic oscillations in different frequency ranges. Just like ERP's, these oscillations can also be analyzed in an event-related fashion (event-related oscillations or ERO's), and it has been convincingly demonstrated that assessing specific frequencies can often provide fruitful insights into the functional cognitive correlates of these signals (Herrmann et al., 2005).

Oscillations are characterized by their amplitude and phase. Whereas the typical amplitude of an EEG oscillation varies between 0 and 10 μV, the cyclic phase ranges from 0 to 2π. These two essential parameters of an oscillation can be determined at any given point in time. Oscillations differ with respect to their degree of phase-locking to an experimental condition (stimulus) and are correspondingly classified as either spontaneous, evoked or induced. While spontaneous activity is completely uncorrelated with the occurrence of a given stimulus, both induced and evoked oscillations are correlated with stimulating events. In particular, induced oscillations are not strictly phase-locked to stimulus onset and thus vary from experimental trial to trial with respect to their onset times and / or phase jitter. Correspondingly, this type of activity is not visible in the averaged ERP due to cancellation and has to be analyzed by applying special methods, which are not in the focus of the present work (for an overview, consider e.g. Herrmann et al., 2005; or Koenig and Pascual-Marqui, 2009). On the contrary, evoked oscillatory activity has the same phase in every stimulus repetition (i.e. it is phase-locked to stimulus onset across experimental trials), sums over trials and is consequently visible in the averaged ERP.

In principle, every signal can be decomposed into sinusoidal oscillations of different frequencies. To extract a specific frequency from EEG data, several methods such as filtering, Fourier transformation and wavelet analysis exist. To evaluate oscillatory activity in the context of the present work, the wavelet analysis approach appears the most advantageous as it not only allows to identify differences between experimental conditions, but also the specific points in time when these differences occur. Very simply expressed, the wavelet transform can be thought of as the envelope

of a bandpass-filtered ERP in a given frequency range. As a consequence, the time course of changing frequency information can be observed. Moreover, the wavelet transform directly yields the amplitude and phase of the signal oscillations in the analyzed frequency band when it is decomposed by using a complex wavelet function.

However, to explain time-frequency decompositions based on complex wavelet transforms in detail lies far beyond the scope of the present thesis. A sophisticated description of this method can be found in Koenig and Pascual-Marqui (2009) or Herrmann et al. (2005).

3. Original Research Articles

3.1. Experiment 1: Processing of voiced and unvoiced acoustic stimuli in musicians

Full Citation:

Ott, C. G., Langer, N., Oechslin, M. S., Meyer, M., and Jäncke, L. (2011). Processing of Voiced and Unvoiced Acoustic Stimuli in Musicians. *Front Psychol* 2 195.

3.1.1. Abstract

Past research has shown that musical training induces changes in the processing of supra-segmental aspects of speech, such as pitch and prosody. The aim of the present study was to determine whether musical expertise also leads to an altered neurophysiological processing of sub-segmental information available in the speech signal, in particular the voice onset time (VOT). Using high-density EEG recordings we analysed the neurophysiological responses to voiced and unvoiced CV syllables and noise analogues in 26 German speaking adult musicians and non-musicians. From the EEG the N1 amplitude of the event-related potential (ERP) and two microstates from the topographical EEG analysis (one around the N1 amplitude and one immediately preceding the N1 microstate) were calculated to the different stimuli. Similar to earlier studies the N1 amplitude was different to voiced and unvoiced stimuli in non-musicians with larger amplitudes to voiced stimuli. The more refined microstate analysis revealed that the microstate within the N1 time window was shorter to unvoiced stimuli in non-musicians. For musicians there was no difference for the N1 amplitudes and the corresponding microstates between voiced and unvoiced stimuli. In addition, there was a longer very early microstate preceding the microstate at the N1 time window to non-speech stimuli only in musicians. Taken together, our findings suggest that musicians process unvoiced stimuli (irrespective whether these stimuli are speech or non-speech stimuli) differently than controls. We propose that musicians utilise the same network to analyse unvoiced stimuli as for the analysis of voiced stimuli. As a further explanation it is also possible that musicians devote more neurophysiological resources into the analysis of unvoiced segments.

frontiers in PSYCHOLOGY

ORIGINAL RESEARCH ARTICLE
published: 06 September 2011
doi: 10.3389/fpsyg.2011.00195

Processing of voiced and unvoiced acoustic stimuli in musicians

Cyrill Guy Martin Ott[1], Nicolas Langer[1], Matthias Oechslin[2,3], Martin Meyer[1] and Lutz Jäncke[1]*

[1] Division Neuropsychology, Psychological Institute, University of Zurich, Zurich, Switzerland
[2] Geneva Neuroscience Center, University of Geneva, Geneva, Switzerland
[3] Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland

Edited by:
Anne-Lise Giraud, INSERM-Ecole Normale Supérieure, France

Reviewed by:
Mitchell Steinschneider, Albert Einstein College of Medicine, USA
Anne Caclin, INSERM, France

****Correspondence:***
Cyrill Guy Martin Ott, Division Neuropsychology, Institute of Psychology, University Zurich, Binzmühlestrasse 14/Box 25, CH-8050 Zürich, Switzerland.
e-mail: c.ott@psychologie.uzh.ch

Past research has shown that musical training induces changes in the processing of supra-segmental aspects of speech, such as pitch and prosody. The aim of the present study was to determine whether musical expertise also leads to an altered neurophysiological processing of sub-segmental information available in the speech signal, in particular the voice-onset-time. Using high-density EEG-recordings we analyzed the neurophysiological responses to voiced and unvoiced consonant-vowel-syllables and noise-analogs in 26 German speaking adult musicians and non-musicians. From the EEG the N1 amplitude of the event-related potential and two microstates from the topographical EEG analysis (one around the N1 amplitude and one immediately preceding the N1 microstate) were calculated to the different stimuli. Similar to earlier studies the N1 amplitude was different to voiced and unvoiced stimuli in non-musicians with larger amplitudes to voiced stimuli. The more refined microstate analysis revealed that the microstate within the N1 time window was shorter to unvoiced stimuli in non-musicians. For musicians there was no difference for the N1 amplitudes and the corresponding microstates between voiced and unvoiced stimuli. In addition, there was a longer very early microstate preceding the microstate at the N1 time window to non-speech stimuli only in musicians. Taken together, our findings suggest that musicians process unvoiced stimuli (irrespective whether these stimuli are speech or non-speech stimuli) differently than controls. We propose that musicians utilize the same network to analyze unvoiced stimuli as for the analysis of voiced stimuli. As a further explanation it is also possible that musicians devote more neurophysiological resources into the analysis of unvoiced segments.

Keywords: musicians, speech, musical expertise, N1, microstates, topographic pattern analysis, transfer effect, experience dependent plasticity

INTRODUCTION

The past 15 years have seen a vast amount of research on music and associated neural processes. One of the most prominent findings is that the intensive practice and training needed to achieve high proficiency in musicianship leads to structural and functional short and long-term alterations in the brain (for a review, see Schlaug, 2001; Münte et al., 2002; Jäncke, 2009). Researchers have therefore become increasingly aware of the value of musicianship as a model for brain plasticity. In view of this, the question arises whether these music-training induced plasticity effects are restricted particularly to structures and functions underlying the perception and production of music or are more widespread in affecting also other domains such as language and speech.

A number of recent studies indicate that the language network is affected by plastic alterations due to musical training (e.g., Thompson et al., 2003; Schön et al., 2004; Moreno and Besson, 2006; Besson et al., 2007; Marques et al., 2007; Parbery-Clark et al., 2009; Gordon et al., 2010; Marie et al., 2010a,b; Colombo et al., 2011). However, most of these studies focused on pitch processing and prosody, nonetheless showing a substantial transfer effect of intensive musical training on auditory language functions. Efficient analysis of spectral acoustic information in the speech signal is indeed crucial for auditory language perception, but the same is true for temporal cues (Shannon et al., 1995; Davis and Johnsrude, 2007). As musical training involves melodic, supra-segmental and subtle timing issues, the processing of rapidly changing temporal cues in the speech signal and the underlying neural substrates may also be altered in musicians. Studies which have examined musicians and non-musicians in terms of their ability to discriminate segmental phonetic cues mainly focused on the spectral domain and generally have revealed negative results. For example, Hillenbrand et al. (1990) have demonstrated that musicians are not faster or even better in discriminating synthetic speech sounds. A similar finding has been provided by Marie et al. (2010a), who examined whether musical expertise has an effect on the discrimination of tonal and segmental variations in tone language. In addition, they measured event-related potentials (ERPs) and identified that musical expertise did not influence segmental processing as indicated by the amplitudes and latencies of early components like the N1. These authors, however, demonstrated strong influences of musical expertise on the later occurring "cognitive" ERP components (N2, P3a, and P3b).

Anyhow, the speech signal carries not only rapidly changing spectral, but also temporal information on a segmental level. Musical training on the other hand, includes practising the perception of rhythm, meter, and subtle timing issues. Thus, it might be the case that when focusing on the temporal domain, musicians and non-musicians actually do differ with respect to the processing of segmental phonetic cues. One of the most important temporal acoustic cues carrying linguistically and phonetically relevant information is the voice-onset-time (VOT). This mainly refers to voicing differences between stop-consonants in a broad range of languages. VOT is defined as the duration of the delay between release of closure and start of voicing (Lisker and Abramson, 1964). Electrophysiological studies have consistently demonstrated VOT-related auditory evoked potential (AEP) differences in the N1 component in humans and animals (e.g., Sharma and Dorman, 1999; Steinschneider et al., 1999, 2005; Sharma et al., 2000; Trébuchon-Da Fonseca et al., 2005; Zaehle et al., 2007). The N1 component is considered to be particularly important in reflecting the basic encoding of acoustic information of the auditory cortex (Näätänen and Picton, 1987; Picton et al., 1987). Moreover, it has been shown that the electrophysiologic patterns elicited by the processing of consonant-vowel (CV) syllables and corresponding non-speech sounds with similar temporal characteristics are comparable in humans (Liégeois-Chauvel et al., 1999; Zaehle et al., 2007). This indicates a general mechanism for encoding and analyzing successive temporal changes in acoustic signals. Modulations of the N1 component reflect the central auditory representation of speech and non-speech sounds (Ostroff et al., 1998; Sharma et al., 2000; Meyer et al., 2006; Pratt et al., 2007). The present study thus aimed to examine potential differences between musicians and non-musicians in the processing of changing temporal acoustic cues. This was done by recording and comparing scalp AEP's in response to CV-syllables and corresponding noise-analogs with varying VOT's and noise-onset times (NOT), respectively. Based on the findings of Zaehle et al. (2007), which clearly demonstrate that the AEP's N1 component is especially responsive to temporal variations in speech and non-speech stimuli, we focused our ("classical") AEP analysis on amplitude modulations at the vertex electrode in the N1 time window. These modulations were elicited by brief auditory stimuli that varied systematically along an acoustic and a linguistic dimension.

We hypothesized that within the non-musicians group we would find stronger N1 amplitudes for stimuli with short (voiced) compared to long (unvoiced) VOT's/NOT's, as this has previously been shown by Zaehle et al. (2007) and various others (e.g., Simos et al., 1998; Sharma et al., 2000; Steinschneider et al., 2005). On the other hand, if musical expertise has an influence on early phonetic processing in the temporal domain, then musicians should show a different pattern in the corresponding neurophysiological indicators of early phonetic processing within the time window of the N1.

In order to reveal potential differences between musicians and non-musicians in distinct stages of information processing of VOT's/NOT's, we also analyzed the spatial variations of the scalp voltage distribution over time between conditions and groups, an approach known as Topographic Pattern Analysis or evoked potential (EP) mapping. This method searches for time segments of stable map topography that represents functional microstates of the brain. These are assumed to reflect distinct information processing steps and provide several advantages over the classical ERP analysis, such as experimenter and reference independence (for an overview, see, e.g., Murray et al., 2008; Michel et al., 2009). This kind of analysis benefits mostly from the high topographic resolution that high-density EEG-recordings provide. Using this approach, we searched for stable map topographies before, during, and after the N1 time window. Here, we are interested in examining whether the duration of these maps are different between musicians and non-musicians.

MATERIALS AND METHODS

SUBJECTS

Thirty-one healthy volunteers with no history of neurological pathology and with normal audiological status participated in this study. All participants were native German or Swiss German speakers. One group consisted of 16 musicians (11 women, 5 men, mean age ± SD of 26.7 ± 10.1 years), with formal training starting at a mean age of 6.02 ± 2.2 years. All musicians were students, music teachers and/or members of an orchestra/band and practiced their instrument daily for 1–6 h. Most of them play several instruments of which the most frequent were piano, violin, and clarinet. The other group comprised 15 non-musicians (11 women, 4 men, mean age ± SD of 26.07 ± 7.9 years) with no formal musical training and no history of musical instrumental performance. According to the Annett-Handedness-Questionnaire (Annett, 1970), 13 participants within the musician group were consistently right-handed, whereas two were ambidextrous and one was left-handed. In the non-musician group, 11 participants were consistent right-handers, one was ambidextrous and three were left-handed. In order to determine each participant's degree of musical competence, the "Advanced Measures of Music Audiation" by Gordon (1989) was applied. All participants performed also a short intelligence test (KAI) to rule out significant differences in intelligence between groups. Descriptive statistics for the criterion measures completed by the participants are summarized in **Table 1**. All participants gave informed written consent, in accordance with procedures approved by the local ethics committee, and were paid for participation.

STIMULI

All participants heard two different classes of auditory stimuli in the context of a phonetic categorization task. One class formed the "speech condition" consisting of the same CV-syllables used in previous studies, such as Jäncke et al. (2002) and Meyer et al. (2007). These syllables (/ka/,/ta/,/pa/,/ga/,/da/,/ba/) were digitally recorded by a trained phonetician at a sampling rate of 44.1 kHz and a sampling depth of 16-bit. The onset, duration, intensity, and fundamental frequency of the stimuli were edited and synchronized by means of a speech editor. The criterion for temporal alignment of the syllables was the onset of articulatory release. The duration of the syllables ranged from 310 to 360 ms, depending on the VOT, with a vowel duration of 300 ms (VOTs in ms for the stops were approximately "k" = 59, "t" = 49, "p" = 39, "g" = 19, "d" = 04, "b" = 09). The second class, forming the "non-speech condition," consisted of noise-analogs, which were synthesized from the CV-syllables used in the speech condition. This was achieved using a variation of a formerly published procedure (Shannon et al., 1995). Thus, spectral information was removed from the syllables by replacing

the frequency specific information in a broad frequency region with band-limited white noise (Band 1: 500–1500 Hz, Band 2: 2500–3500 Hz) whereas normalized amplitudes and temporal cues were preserved in each spectral band, resulting in double-band-pass filtered noise with temporal CV-amplitude dynamics. Thus, these stimuli are devoid of any phonological, semantic, and syntactic cues but imitate the temporal envelope of proper speech signals (syllables, words, or sentences).

This non-speech condition was introduced in a non-task-relevant manner to reassess whether any found differences are due to different VOT processing and not to other varying characteristics between the speech stimuli. **Figure 1** shows both the speech and non-speech waveforms of the syllables "pa" and "ba" as an example of the stimuli used in the present study. All manipulations were carried out with the "Adobe Audition 1.5" software (Adobe Systems Incorporated, 1992–2004). All auditory stimuli were presented binaurally at a sound pressure level of about 70 dB using hi-fi headphones. Stimuli were presented and responses recorded using Presentation® software (Neurobehavioral Systems, USA).

PROCEDURE

The EEG experiment comprised four blocks per condition (i.e., "speech" and "non-speech"), resulting in a total of eight blocks and an overall duration of the experiment of about 60 min. Each block contained 42 trials of each stimulus, thus presenting 252 stimuli per block and a total amount of 168 presentations per stimulus. In both conditions, stimuli were shuffled in a pseudo-randomized order. The condition order was randomized for every subject. Stimuli were of two different categories, assigned either "voiced" or "unvoiced," depending on their particular VOT ("voiced" = /ga/,/da/,/ba/and corresponding noise-analogs; "unvoiced" = /ka/,/ta/,/pa/and corresponding noise-analogs). Participants were seated in a comfortable chair in an electrically shielded, dimly lit room, and were instructed to listen attentively to the auditory stimuli presented via headphones. They had to decide as quickly and accurately as possible whether the heard stimulus was of type "voiced" or "unvoiced," and to indicate their response by pressing either the left or right mouse button. We utilized this task to ensure the participant's vigilance throughout the

Table 1 | Descriptive statistics for the criterion measures completed by the subjects are listed group-wise (M = musicians; NM = non-musicians; GMA = Gordon musical aptitude).

	Age		Gender		IQ		GMA*		Handedness		
	Mean	SD	m	f	Mean	SD	Mean	SD	Left	Right	Ambi
M	26.7	10.1	5	11	131.4	6.5	68.93	14.28	1	13	2
NM	26.07	7.9	4	11	124.9	11.9	43.26	21.41	3	11	1

**Difference between experimental groups is significant at $p < 0.05$ [$T(29) = -3.951$; note that t-tests were only applied to variables "Age," "IQ" and "GMA." For "Gender" and "Handedness," Chi-Square tests were used]. All statistical tests shown in this table were conducted comparing the two experimental groups.*

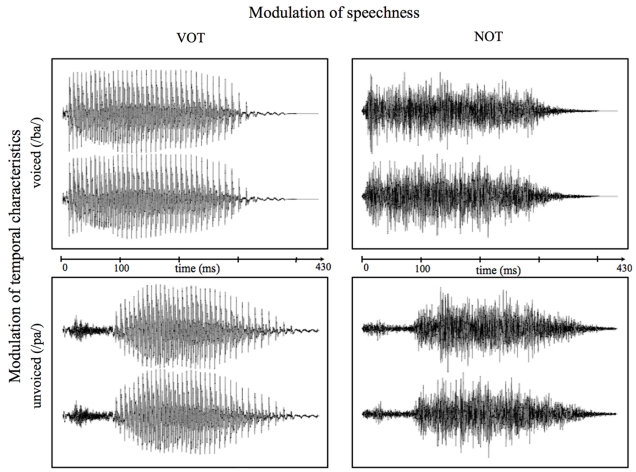

FIGURE 1 | Exemplary waveforms of voiced (/ba/; top) and unvoiced (/pa/; bottom) speech (left) and correspondent non-speech (right) stimuli. Time point "0" refers to the stimulus onset.

experiment, even though it has previously been shown (Baumann et al., 2008) that an enhancement of N1 amplitudes in musicians rather reflects an influence of expertise than selective attention. In order to ensure that participants could solve the task as instructed, each participant performed practice trials before the experiment until complete comprehension of the task was reported. As each stimulus was presented 500 ms after the preceding mouse button press, the inter-stimulus interval depended mainly on the speed of the participant's response. When no response occurred, the next stimulus was presented automatically after 2 s.

AEP RECORDINGS

Electroencephalogram was recorded using a high-density Geodesic EEG system® (GSN300; Electrical Geodesic Inc., Oregon)[1] with 128-Channel HydroCel Geodesic Sensor Nets@ (HCGSN120). Data was sampled at 250 Hz and band-pass filtered at 0.1–100 Hz. The vertex electrode (Cz) served as on-line recording reference. Impedance was kept below 30 kΩ. For various reasons, such as violations of attendance criteria regarding age or technical problems during EEG recording, behavioral and EEG data of five participants (three musicians and two non-musicians) were discarded from further analysis, leading to a total of 26 remaining datasets (13 per group).

DATA ANALYSIS – BEHAVIORAL DATA

The number of correct identifications, errors, and misses in each block was counted within the two groups (musicians vs. non-musicians) and averaged for each subject across blocks, categories (i.e., "voiced" vs. "unvoiced"), and classes (i.e., "speech" vs. "non-speech"). The same procedure was carried out with reaction times (RT), which were analyzed for correct identifications only. Thus, a $2 \times 2 \times 2$ repeated measure analysis of variance (ANOVA) with factors *group* (between-subject), *voiceness*, and *speechness* (within-subject) was computed for each of these variables in order to statistically evaluate any differences in identification accuracy and RT's.

DATA ANALYSIS – EEG, N1 COMPONENT ANALYSIS

Electroencephalogram recordings of each participant were imported and analyzed in the BrainVision Analyzer2 software (Version 2.0.1; Brain Products GmbH, D-82205 Gilching)[2]. In a first step, data was band-pass filtered at 1.5–70 Hz and a 50-Hz notch filter was applied. An ICA (independent component analysis) was then performed to correct for ocular artifacts (e.g., Jung et al., 2000). Each EEG recording was visually inspected and trials with sweating and muscle artifacts, amplifier saturation, and remaining eye blinks or eye movements were excluded. Noisy channels were interpolated and the data was then re-referenced to the mastoid electrodes for ERP calculation. Each ERP waveform was an average of more than 100 repetitions of the EEG sweeps evoked by the same stimulus type. EEG-recordings were sectioned into 500 ms segments (100 ms pre-stimulus and 400 ms post-stimulus) and a baseline correction using the signal's pre-stimulus portion was carried out. Finally, ERP's for each stimulus were averaged for each subject and grand-averaged across subjects within the two groups separately. Had peak values of grand-averages been used for statistical analysis, differences between experimental groups in the N1 peak values might have been blurred during averaging due to individual differences in their latencies. Therefore, grand-average waves were used only to reassess the N1 time window and for illustrative purposes.

In order to statistically confirm relevant differences between AEP's at Cz as a function of experimental group, stimulus category and condition, mean amplitude ERP's time-locked to the auditory stimulation were therefore measured in a specific latency window. This was individually pre-determined for each subject and stimulus by visual inspection of the event-related signal. These individual latency windows were centered at the peak of the prominent N1 component and covered a total signal length of 20 ms around the center. Individually chosen latency windows were used to ensure that the mean amplitude actually reflect the N1 peak values of every subject and stimulus. Mean amplitudes were then averaged separately within groups (i.e., "musicians" vs. "non-musicians"), depending on their category (i.e., "voiced" vs. "unvoiced"), and class (i.e., "speech" vs. "non-speech"). Subsequently, a $2 \times 2 \times 2$ repeated measure ANOVA with a between-subject factor (*group*) and two within-subjects factors (*speechness* and *voiceness*) was computed for the central electrode (Cz), and the reported p-values were adjusted for non-sphericity using Greenhouse–Geisser Epsilon when equal variances could not be assumed. Subsequently, Bonferroni–Holm adjusted *post hoc t*-tests were applied. The global level of significance was set at $p < 0.05$ for all statistical analyses.

DATA ANALYSIS – EEG, TOPOGRAPHIC PATTERN ANALYSIS

Topographic Pattern Analysis was performed using the Cartool software (Version 3.43/869; The Cartool Community group)[3]. Subject- and grand-averaged ERP's were imported into Cartool and recalculated against the average reference. To identify the dominant map topographies on the scalp, segments of stable voltage topography (or EP maps) were then defined by using a topographic atomize and agglomerate hierarchical cluster analysis (T-AAHC) in the grand-averaged ERP's across conditions and groups over the full post-stimulus EEG segment length of 400 ms. These template maps are the mean maps over the period where the stable voltage topography segments were found. The optimal number of templates was determined by the combination of a modified cross-validation and the Krzanowski–Lai criterion (e.g., Pascual-Marqui et al., 1995; Murray et al., 2008). In order to assess any differences regarding map duration between groups and conditions in and around the N1 time window (i.e., 50–200 ms), the presence of each map that was identified in the group-averaged data within that particular epoch was verified statistically over the same period in the ERP's of the individual participants (i.e., "single-subject fitting"; Murray et al., 2008). This step allowed us to determine the duration of any given template map for each condition within the musician's and control's group between 50 and 200 ms. This time window was accordingly chosen to encompass the N1 time window itself plus the periods immediately before and after it. These duration values were then statistically evaluated

[1] http://www.egi.com
[2] http://www.brainproducts.com
[3] sites.google.com/site/cartoolcommunity

for each map of interest by means of a repeated measures ANOVA with the factors *group* (between-subjects), *speechness,* and *voiceness* (within-subjects), as in the classical N1 component analysis. Subsequently, any significant interactions were further examined by applying *post hoc t*-tests.

RESULTS

BEHAVIORAL DATA

We analyzed accuracy and RT using 2 × 2 × 2 ANOVAs with one between-subjects factor (*group:* musicians vs. non-musicians) and two repeated measurement factors (*voiceness:* voiced vs. unvoiced; *speechness:* speech vs. non-speech). With regard to accuracy, significant main effects for *voiceness* [$F(1, 24) = 6.096$, $p < 0.05$] and *speechness* [$F(1, 24) = 6.742$, $p < 0.05$] were found with higher accuracies for speech stimuli in general. Furthermore, there was a significant *voiceness* by *speechness* [$F(1,24) = 1.217$, $p < 0.05$] interaction characterized by higher accuracies for voiced than unvoiced non-speech stimuli [$T(25) = 5.296$, $p < 0.05$] and for unvoiced speech than unvoiced non-speech stimuli [$T(25) = 3.929$, $p < 0.05$]. There were no significant differences between groups neither for the main effects nor the interactions. With respect to RT, neither significant main nor interaction effects were found in the 2 × 2 × 2 ANOVA of speech and non-speech stimuli. Mean reaction times for the two experimental groups are shown in **Figure 2A**. **Figure 2B** shows overall accuracies irrespective of group. Note that only RT of correct answers were analyzed.

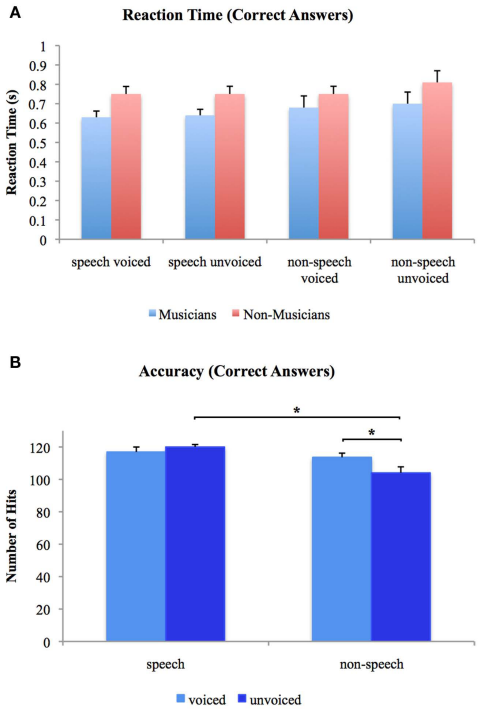

FIGURE 2 | (A) Reaction times to voiced and unvoiced speech and non-speech stimuli for musicians and non-musicians. **(B)** Overall accuracies for voiced and unvoiced speech and non-speech stimuli. *Differences are significant at $p < 0.05$ [differences between stimulus class (i.e., "speech" vs. "non-speech") and stimulus type (i.e., "voiced" vs. "unvoiced") irrespective of experimental group (i.e., "musicians" vs. "non-musicians"), $T(25) = 5.296$ and $T(25) = 3.929$, respectively]. Error bars indicate standard errors.

EEG DATA – N1 COMPONENT ANALYSIS

Grand-averaged waveforms evoked by each category and class of stimuli recorded at Cz are shown in **Figure 3** for the speech and in **Figure 4** for the non-speech condition, respectively. We observed that all stimuli elicited a prominent N1 component. In the non-musician group, the strongest N1 amplitudes were found for stimuli with short VOT's (i.e., "voiced" stimuli), thus replicating the results of Zaehle et al. (2007) and various others (e.g., Simos et al., 1998; Sharma et al., 2000). However, in the musician group, no difference between N1 amplitudes was revealed. In order to statistically examine the ERP effects, mean amplitudes of each stimulus category (i.e., "voiced" vs. "unvoiced") and class (i.e., "speech" vs. "non-speech") were analyzed for each subject in a 20-ms latency window around the N1 peaks.

Results of the 2 × 2 × 2 repeated measures ANOVA with the factors *group* (musicians vs. non-musicians), *voiceness* (voiced vs. unvoiced), and *speechness* (speech vs. non-speech) for the N1 amplitudes showed significant main effects for the factors *voiceness* [$F(1, 24) = 4.733$, $p < 0.05$] and *speechness* [$F(1, 24) = 29.33$, $p < 0.05$]. In addition, a significant *group* by *voiceness* interaction was revealed [$F(1, 24) = 6.209$, $p < 0.05$].

Separate *post hoc* analyses were conducted for the two experimental groups, using Bonferroni–Holm adjusted *t*-tests and accepting therefore differences as significant only when the corresponding *p*-values were below the adjusted significance-level. **Figures 3 and 4** show group-wise plots of the mean amplitudes recorded at Cz for each stimulus category. **Figure 5** shows the corresponding results of *post hoc t*-tests collapsed for both the speech and non-speech condition (i.e., speech and non-speech stimuli are treated as one class). In the non-musician group, separate *t*-tests revealed significantly stronger N1 potentials elicited by voiced than by unvoiced stimuli [$T(12) = -3.480$, $p < 0.05$]. In contrast, neither significant differences were found within the musician group nor between groups. Moreover, N1 amplitudes elicited by speech were stronger in general than those elicited by non-speech [$T(25) = 5.446$, $p < 0.05$].

EEG DATA – TOPOGRAPHIC PATTERN ANALYSIS

Figure 6 shows the results of the topographical EP mapping of the grand-averaged data for each group and condition. In order to provide consistency with the examination of the classical N1 component, we first inspected visually the resulting template maps in and around the N1 time window (50–200 ms). Subsequently, we restricted our further analyses to the one template map, which expressed a typical N1-like auditory topography (map 6) and those which appeared immediately before and after that map (maps 3 and 7). Durations of these maps were then extracted and compared by means of a repeated measure ANOVA.

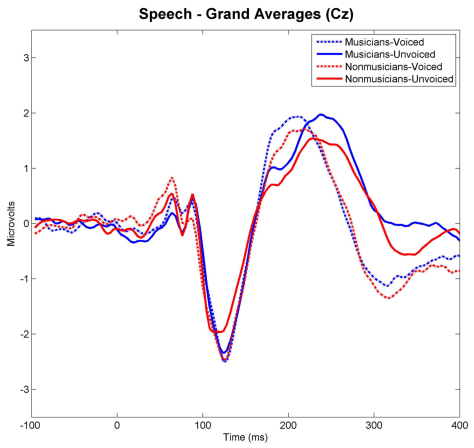

FIGURE 3 | Grand-averaged waveforms elicited by voiced and unvoiced speech stimuli. Time point "0" refers to the stimulus onset.

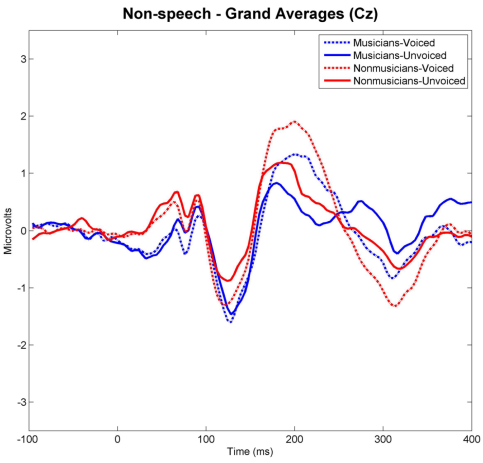

FIGURE 4 | Grand-averaged waveforms elicited by voiced and unvoiced non-speech stimuli. Time point "0" refers to the stimulus onset.

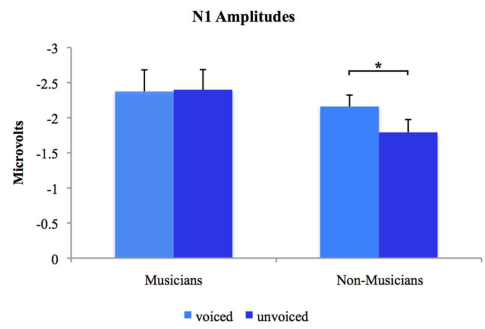

FIGURE 5 | Group-wise plots of N1 mean amplitudes for voiced and unvoiced stimuli. Speech and non-speech conditions are collapsed. *Difference is significant at $p < 0.05$ [difference between stimulus type (i.e., "voiced" vs. "unvoiced") irrespective of stimulus class (i.e., "speech" vs. "non-speech") within the non-musician group, $T(12) = -3.480$]. Error bars indicate standard errors.

Results of the 2 × 2 × 2 repeated measure ANOVA with between-subjects factor *group* and within-subjects factors *speechness* and *voiceness* showed significant main effects for the factor *speechness* regarding the durations of map 3 [$F(1, 24) = 10.046$, $p < 0.05$] and map 7 [$F(1, 24) = 5.139$, $p < 0.05$]. Furthermore, a significant *speechness by group* interaction was found for map 3 [$F(1, 24) = 6.684$, $p < 0.05$], whereas map 6 exhibited a significant *voiceness by group* interaction [$F(1, 24) = 5.933$, $p < 0.05$].

Subsequent *post hoc* Bonferroni–Holm adjusted *t*-tests revealed significant differences regarding the duration of map 3 between speech and non-speech stimuli [$T(12) = -3.582$, $p < 0.05$] within the musician group. In particular, map 3 was longer for musicians in the non-speech condition. Non-musicians on the other hand showed similar durations for map 3 in both conditions.

The topography of map 6 strongly resembles the typical auditory N1 topography. *Post hoc t*-tests for the durations of this microstate revealed significantly longer durations for voiced than for unvoiced stimuli within the non-musicians group [$T(12) = 2.533$, $p < 0.05$]. Within the musicians group, no significant difference in the duration of map 6 was found at all.

DISCUSSION

The purpose of this study was to examine the extent to which musicians and non-musicians process phonetic information differently. We were specifically interested in the early processing steps within the first 200 ms of stimulus onset during which the basic acoustic properties of acoustic stimuli are processed. Our hypothesis was that musical practice would have shaped the auditory system of musicians and that this use-dependent influence might also exert its influence on early phoneme perception. We anticipated therefore different neurophysiological activation patterns during early auditory processing stages.

For this study we used speech and non-speech stimuli with short (voiced) and long (voiceless) VOT. High-density EEG-recordings were obtained during the early acoustic processing stages and analyzed in two different ways: Firstly we calculated and analyzed conventional AEP components and focused on the N1 component, which is particularly important for auditory analysis (e.g., Sharma and Dorman, 1999; Steinschneider et al., 1999, 2005; Sharma et al., 2000; Zaehle et al., 2007). In addition to the conventional ERP analysis we also applied a topographic pattern analysis. With this method we exploit the entire spatial information of our high-density EEG

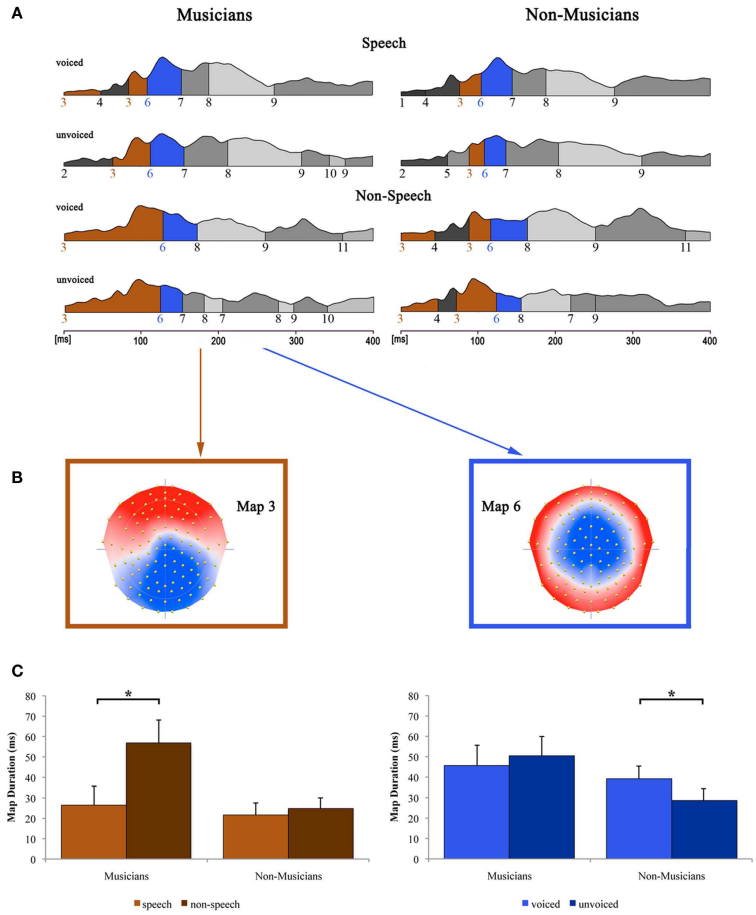

FIGURE 6 | Evoked potential data. (A) Segments of stable map topography are shown group-wise (left: musicians, right: non-musicians) for each condition under the global field power curve from 0 to 400 ms. Auditory N1-like template map 6 (blue) was found at ~100 ms and was significantly longer for voiced than for unvoiced stimuli within the non-musicians group whereas within the musicians group, durations were comparable. Template map 3 (brown) preceded map 6 and lasted significantly longer in the non-speech condition compared to the speech condition within the musicians group only. **(B)** Map topography and **(C)** duration of maps 3 (left) and 6 (right) for musicians and non-musicians in the respective conditions (voiced and unvoiced stimuli are collapsed for map 3; speech and non-speech conditions are collapsed for map 6). *Difference is significant at $p < 0.05$ [left: Difference between stimulus class (i.e., "speech" vs. "non-speech") irrespective of stimulus type (i.e., "voiced" vs. "unvoiced") within the musician group, $T(12) = -3.582$; right: Difference between stimulus type irrespective of stimulus class within the non-musician group, $T(12) = 2.533$]. Note that the data shown in **(A,B)** are derived of grand-averaged waveforms, whereas map durations shown in **(C)** are obtained from individual ERPs (for a detailed description of the microstate analysis procedure see, e.g., Murray et al., 2008). Error bars indicate standard errors.

recording and reveal a more detailed activation pattern. Using this approach we focused on microstates around the N1 time window. Microstates are identified in an entirely data-driven approach, which substantially decreases the subjective influence on the data analysis. Thus, this method can be considered a more objective method to analyze evoked electrical responses (Michel et al., 2009).

As in various aforementioned studies (e.g., Sharma et al., 2000; Zaehle et al., 2007), the N1 amplitude is different to voiced and unvoiced acoustic stimuli in non-musicians. In particular, the N1 is larger for voiced than for unvoiced stimuli. Musicians on the other hand demonstrated large N1 amplitudes to all stimuli with no difference between voiced and unvoiced stimuli. Correspondingly,

for non-musicians the microstate durations around the N1 time window are significantly shorter for unvoiced stimuli irrespective of whether these stimuli are "speech" or "non-speech." In addition, there is no difference between the microstate durations for musicians.

The findings with respect to the N1 amplitude and the corresponding microstate supports the idea that non-musicians process voiced and unvoiced stimuli differently and utilize different neurophysiological processes at least at processing stages around the N1 component. Simos et al. (1998) found the N1m peak equivalent current dipole (ECD) to be shifted toward more medial locations with increasing VOT. Due to the authors, this finding suggests that the peak of the N1m response elicited by stimuli with short VOT values may reflect activity from distinct subpopulations of auditory cortex neurons compared to the N1m response elicited by stimuli with long VOT values. That different neurophysiological processes are used to analyze voiced and unvoiced stimuli reflects the fact that there are qualitative differences between voiced and unvoiced phonemes. A pivotal feature in the perception of unvoiced phonemes is the relatively long VOT and the characteristic VOT differences between different unvoiced phonemes. For example, the VOT for/k/is much longer than the one for/p/. The VOT of voiced phonemes on the other hand is fairly short in general and much harder to resolve. Therefore, the processing of other acoustic features such as formant transitions and the specific composition of formants becomes more important in order to perceive and differentiate voiced phonemes, and spectral information is more relevant for this kind of analysis. Voiced phonemes are harder to detect in monitoring tasks (not in our study since the task used in this experiment was fairly easy) resulting in reduced hit rates and longer RT (Jäncke et al., 1992), thus it is possible that the processing of voiced stimuli requires more neurophysiological resources indicated by larger N1 amplitudes. Analysis of VOTs for perceiving unvoiced stimuli is a bit easier and thus does not need that much neurophysiological resource. A further possibility could be that the long VOT in unvoiced syllables provides the opportunity for some pre-processing after the first consonant, which then limits the quantity of processing needed after the occurrence of the following vowel. This preparation (or "warning effect") elicited by the long VOT could explain the reduced N1 for unvoiced stimuli. If this argument is correct then musicians would show either less of this "preparation effect" or their auditory system would recover more quickly (or efficiently) after the "warning" initial consonant than in non-musicians. Finally, N1 amplitudes in response to unvoiced stimuli could also be reduced due to a second positive-going wave elicited by voicing onset that truncates the N1 evoked by consonant release (e.g., Sharma et al., 2000; Steinschneider et al., 2005).

Taken together, non-musicians utilize different processing modes at the time window of the N1 component for voiced and unvoiced stimuli.

The major finding of our study is that musicians process unvoiced stimuli differently than non-musicians. The N1 amplitude and the corresponding microstate are practically identical for voiced and unvoiced stimuli in musicians. Does that mean that musicians process voiced and unvoiced stimuli similarly? Although we are not in the position to answer this question unequivocally on the basis of our data, we believe that musicians indeed process voiced and unvoiced stimuli at least partly similar. Due to their lifelong musical practise to perceive, distinguish and memorize pitch and supra-segmental information, they are experts in this regard. In this context, the primary and secondary auditory cortex has been adapted anatomically and neurophysiologically to effectively analyze pitch and supra-segmental information (e.g., Pantev et al., 1998, 2003; Schneider et al., 2002; Baumann et al., 2008; Brancucci et al., 2009; Geiser et al., 2009). This expertise can even be identified when the musicians pre-attentively process pitch and supra-segmental information as has been shown in many studies using the mismatch negativity potential (e.g., Pantev et al., 2003; Herholz et al., 2008, 2009; Chandrasekaran et al., 2009; Nikjeh et al., 2009). Because of this preponderance, it might be that musicians are prone to use the neurophysiological processes applied for pitch-relevant information also to the analysis of unvoiced stimuli. Thus, they activate the pitch information analysis mode (*voiced analysis mode*) even during analyses of unvoiced stimuli. A further possibility could be that they activate the two analysis modes simultaneously when analyzing unvoiced stimuli, namely the *voiced* and the *unvoiced mode*. This would imply that musicians run a kind of multi-tasking analysis, which possibly provides more refined information about the spectro-temporal features of an auditory event. It might also be the case that highly trained musicians invest processing resources to a higher extent in the analysis of unvoiced stimuli, in order to ensure the analysis to be as precise as possible. Whether these explanations hold true has to be shown in future experiments.

Map 3, a microstate which immediately precedes the microstate within the N1 time window (here map 6) also revealed differences between musicians and non-musicians. This microstate lies in the P1 AEP time window (50–100 ms) and reflects early auditory neurophysiological processes. For this microstate musicians demonstrate longer durations during processing of non-speech stimuli irrespective of whether these stimuli are voiced or unvoiced. Thus, this neurophysiological process is longer active in musicians than in non-musicians. Even though our experiment was not primarily focused on this particular time window and we thus had no *a priori* hypothesis in this respect, we explain this finding as follows: We believe that in musicians, non-speech stimuli automatically evoke the *non-speech processing mode* at this very early processing stage. This processing mode is most likely specific for musicians because they have practised a lot to improve their capabilities to analyze speech-unspecific acoustic features (e.g., pitch, rhythm, intensity, metrum). Thus, they remain longer in the processing stage during which these basic acoustic features are analyzed than non-musicians.

Beside these neurophysiological differences there was also a slight advantage with respect to the RT for musicians. Though not significant, this was a consistent effect, especially for the speech stimuli. Whether this behavioral advantage is related to superior auditory processing or due to a more efficient auditory-motor coupling (see for example Bangert and Altenmüller, 2003) is difficult to infer. Future studies need to be conducted using different experimental designs in order to disentangle auditory processing from audio-motor coupling.

Most of the studies published so far which have examined differences between musicians and non-musicians with respect to basic auditory analyses have focused on supra-segmental acoustic features and pitch processing (e.g., Pantev et al., 1998, 2003; Schneider et al.,

2002; Magne et al., 2003; Moreno and Besson, 2005, 2006; Chartrand and Belin, 2006; Baumann et al., 2008; Brancucci et al., 2009; Geiser et al., 2009; Oechslin et al., 2009). Others examined short-term effects of musical training on brain plasticity and its behavioral outcome with respect to speech (e.g., Magne et al., 2006; Moreno et al., 2008; Draganova et al., 2009; Nikjeh et al., 2009). They clearly show that the auditory cortex is functionally altered in musicians compared to non-musicians, thus reflecting their exceptional expertise in encoding and analyzing auditory events. Moreover, structural and functional alterations can be seen already after a relatively short exposure to musical training. The present study complements these findings insofar as the results at hand show also different processing of voiced and unvoiced phonemes on a segmental level. This is reflected by different activation patterns between musicians and non-musicians with respect to N1 amplitudes and corresponding map durations.

It is however important to point out that it is not clear yet if and to what extent these differences might be driven by functionally altered processes at an even earlier level. Musacchia et al. (2007) clearly demonstrated that musical expertise affects the processing of speech as well as music at the level of the brainstem. In particular, they found stronger amplitudes in musicians to both CV-syllables and musical sounds in the transient and sustained portions of the brainstem response to auditory events. Several further studies examining the frequency-following response (FFR), which is interpreted as a marker of brainstem activity, have produced results in line with these findings. For example, Bidelman et al. (2011) found stronger subcortical pitch encoding in musicians compared to non-musicians as reflected in an enhanced FFR in musicians for both speech and non-speech stimuli. It is therefore conceivable that the results of the present study regarding functional differences at the cortical level are influenced by or even solely based on an already altered afferent signal from the brainstem.

CONCLUSION

Our findings indicate that musical expertise transfers to basal speech processing on a very basic level of processing. Musicians utilize a similar neurophysiological processing mode to analyze voiced and unvoiced stimuli at the time window of the N1 component. Maybe they activate both modes (voiced and unvoiced analyses) simultaneously. At an earlier processing step preceding the voiced–unvoiced processing mode they strongly react to non-speech stimuli indicating a kind of automatic processing mode for non-speech stimuli. Although there are substantial neurophysiological differences between musicians and non-musicians, there are no strong behavioral differences. However, our results provide further support for the view that language and music share neural resources.

ACKNOWLEDGMENT

This study was supported by a grant of the Swiss National Science Foundation (SNF) to Lutz Jäncke (32003B_121927).

REFERENCES

Annett, M. (1970). A classification of hand preference by association analysis. *Br. J. Psychol.* 61, 303–321.

Bangert, M., and Altenmüller, E. O. (2003). Mapping perception to action in piano practice: a longitudinal DC-EEG study. *BMC Neurosci.* 4, 26. doi: 10.1186/1471-2202-4-26

Baumann, S., Meyer, M., and Jäncke, L. (2008). Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *J. Cogn. Neurosci.* 20, 2238–2249.

Besson, M., Schön, D., Moreno, S., Santos, A., and Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restor. Neurol. Neurosci.* 25, 399–410.

Bidelman, G. M., Gandour, J. T., and Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J. Cogn. Neurosci.* 23, 425–434.

Brancucci, A., di Nuzzo, M., and Tommasi, L. (2009). Opposite hemispheric asymmetries for pitch identification in absolute pitch and non-absolute pitch musicians. *Neuropsychologia* 47, 2937–2941.

Chandrasekaran, B., Krishnan, A., and Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain Lang.* 108, 1–9.

Chartrand, J. P., and Belin, P. (2006). Superior voice timbre processing in musicians. *Neurosci. Lett.* 405, 164–167.

Colombo, L., Deguchi, C., Boureux, M., Sarlo, M., and Besson, M. (2011). Detection of pitch violations depends upon the familiarity of intonational contour of sentences. *Cortex* 47, 557–568.

Davis, M. H., and Johnsrude, I. S. (2007). Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hear. Res.* 229, 132–147.

Draganova, R., Wollbrink, A., Schulz, M., Okamoto, H., and Pantev, C. (2009). Modulation of auditory evoked responses to spectral and temporal changes by behavioral discrimination training. *BMC Neurosci.* 10, 143. doi: 10.1186/1471-2202-10-143

Geiser, E., Ziegler, E., Jancke, L., and Meyer, M. (2009). Early electrophysiological correlates of meter and rhythm processing in music perception. *Cortex* 45, 93–102.

Gordon, E. E. (1989). *Advanced Measures of Music Audiation*. Chicago, IL: GIA Publications, Inc.

Gordon, R. L., Schön, D., Magne, C., Astesano, C., and Besson, M. (2010). Words and melody are intertwined in perception of sung words: EEG and behavioral evidence. *PLoSONE* 5e, 9889. doi: 10.1371/journal.pone.0009889

Herholz, S. C., Lappe, C., Knief, A., and Pantev, C. (2008). Neural basis of music imagery and the effect of musical expertise. *Eur. J. Neurosci.* 28, 2352–2360.

Herholz, S. C., Lappe, C., and Pantev, C. (2009). Looking for a pattern: an MEG study on the abstract mismatch negativity in musicians and nonmusicians. *BMC Neurosci.* 10, 42. doi: 10.1186/1471-2202-10-42

Hillenbrand, J., Canter, G. J., and Smith, B. L. (1990). Perception of intraphonemic differences by phoneticians, musicians, and inexperienced listeners. *J. Acoust. Soc. Am.* 88, 655–662.

Jäncke, L. (2009). The plastic human brain. *Restor. Neurol. Neurosci.* 27, 521–538.

Jäncke, L., Steinmetz, H., and Volkmann, J. (1992). Dichotic listening: what does it measure? *Neuropsychologia* 30, 941–950.

Jäncke, L., Wüstenberg, T., Scheich, H., and Heinze, H. J. (2002). Phonetic perception and the temporal cortex. *Neuroimage* 15, 733–746.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., and Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* 37, 163–178.

Liégeois-Chauvel, C., de Graaf, J. B., Laguitton, V., and Chauvel, P. (1999). Specialization of left auditory cortex for speech perception in man depends on temporal coding. *Cereb. Cortex* 9, 484–496.

Lisker, L., and Abramson, A. S. (1964). Across language study of voicing in initial stops: acoustical measurements. *Word* 20, 384–411.

Magne, C., Schön, D., and Besson, M. (2003). Prosodic and melodic processing in adults and children. behavioral and electrophysiologic approaches. *Ann. N. Y. Acad. Sci.* 999, 461–476.

Magne, C., Schön, D., and Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *J. Cogn. Neurosci.* 18, 199–211.

Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., and Besson, M. (2010a). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *J. Cogn. Neurosci.* doi: 10.1162/jocn.2010.21585. [Epub ahead of print].

Marie, C., Kujala, T., and Besson, M. (2010b). Musical and linguistic expertise influence pre-attentive and attentive processing of non-speech

sounds. *Cortex*. doi:10.1016/j.cortex.2010.11.006. [Epub ahead of print].

Marques, C., Moreno, S., Castro, S. L., and Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. *J. Cogn. Neurosci.* 19, 1453–1463.

Meyer, M., Baumann, S., and Jancke, L. (2006). Electrical brain imaging reveals spatio-temporal dynamics of timbre perception in humans. *Neuroimage* 32, 1510–1523.

Meyer, M., Elmer, S., Baumann, S., and Jancke, L. (2007). Short-term plasticity in the auditory system: differential neural responses to perception and imagery of speech and music. *Restor. Neurol. Neurosci.* 25, 411–431.

Michel, C. M., Th. Koenig, Brandeis, D., Gianotti, L. R. R., and Wackermann, J. (2009). "Electrical neuroimaging in the time domain - microstate analysis," in *Electrical Neuroimaging*, Chap. 6, eds C. M., Michel, Th. Koenig, D. Brandeis, L. R. R. Gianotti, and J. Wackermann (Cambridge: Cambridge University Press), 123–125.

Moreno, S., and Besson, M. (2005). Influence of musical training on pitch processing: event-related brain potential studies of adults and children. *Ann. N. Y. Acad. Sci.* 1060, 93–97.

Moreno, S., and Besson, M. (2006). Musical training and language-related brain electrical activity in children. *Psychophysiology* 43, 287–291.

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S., and Besson, M. (2008). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cereb. Cortex* 19, 712–723.

Münte, T. F., Altenmüller, E., and Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nat. Rev. Neurosci.* 3, 473–478.

Murray, M. M., Brunet, D., and Michel, C. M. (2008). Topographic ERP analyses: a step-by-step tutorial review. *Brain Topogr.* 20, 249–264.

Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15894–15898.

Näätänen, R., and Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24, 375–425.

Nikjeh, D. A., Lister, J. J., and Frisch, S. A. (2009). Preattentive cortical-evoked responses to pure tones, harmonic tones, and speech: influence of music training. *Ear Hear.* 30, 432–446.

Oechslin, M. S., Meyer, M., and Jäncke, L. (2009). Absolute pitch-functional evidence of speech-relevant auditory acuity. *Cereb. Cortex* 20, 447–455.

Ostroff, J. M., Martin, B. A., and Boothroyd, A. (1998). Cortical evoked response to acoustic change within a syllable. *Ear Hear.* 19, 290–297.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., and Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature* 392, 811–814.

Pantev, C., Ross, B., Fujioka, T., Trainor, L. J., Schulte, M., and Schulz, M. (2003). Music and learning-induced cortical plasticity. *Ann. N. Y. Acad. Sci.* 999, 438–450.

Parbery-Clark, A., Skoe, E., Lam, C., and Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear Hear.* 30, 653–661.

Pascual-Marqui, R. D., Michel, C. M., and Lehmann, D. (1995). Segmentation of brain electrical activity into microstates: model estimation and validation. *IEEE Trans. Biomed. Eng.* 42, 658–665.

Picton, T. W., Skinner, C. R., Champagne, S. C., Kellett, A. J., and Maiste, A. C. (1987). Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone. *J. Acoust. Soc. Am.* 82, 165–178.

Pratt, H., Starr, A., Michalewski, H. J., Bleich, N., and Mittelman, N. (2007). The N1 complex to gaps in noise: effects of preceding noise duration and intensity. *Clin. Neurophysiol.* 118, 1078–1087.

Schlaug, G. (2001). The brain of musicians. A model for functional and structural adaptation. *Ann. N. Y. Acad. Sci.* 930, 281–299.

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., and Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat. Neurosci.* 5, 688–694.

Schön, D., Magne, C., and Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* 41, 341–349.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science* 270, 303–304.

Sharma, A., and Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *J. Acoust. Soc. Am.* 106, 1078–1083.

Sharma, A., Marsh, C. M., and Dorman, M. F. (2000). Relationship between N1 evoked potential morphology and the perception of voicing. *J. Acoust. Soc. Am.* 108, 3030–3035.

Simos, P. G., Diehl, R. L., Breier, J. I., Molis, M. R., Zouridakis, G., and Papanicolaou, A. C. (1998). MEG correlates of categorical perception of a voice onset time continuum in humans. *Cogn. Brain Res.* 7, 215–219.

Steinschneider, M., Volkov, I. O., Fishman, Y. I., Oya, H., Arezzo, J. C., and Howard, M. A. (2005). Intracortical response in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. *Cereb. Cortex* 15, 170–186.

Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., and Howard, M. A. (1999). Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *J. Neurophysiol.* 82, 2346–2357.

Thompson, W. F., Schellenberg, E. G., and Husain, G. (2003). Perceiving prosody in speech. effects of music lessons. *Ann. N. Y. Acad. Sci.* 999, 530–532.

Trébuchon-Da Fonseca, A., Giraud, K., Badier, J., Chauvel, P., and Liégeois-Chauvel, C. (2005). Hemispheric lateralization of voice onset time (VOT) comparison between depth and scalp EEG recordings. *Neuroimage* 27, 1–14.

Zaehle, T., Jancke, L., and Meyer, M. (2007). Electrical brain imaging evidences left auditory cortex involvement in speech and non-speech discrimination based on temporal features. *Behav. Brain Funct.* 3, 63.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 21 March 2011; accepted: 03 August 2011; published online: 06 September 2011.

Citation: Ott CGM, Langer N, Oechslin M, Meyer M and Jäncke L. (2011) Processing of voiced and unvoiced acoustic stimuli in musicians. Front. Psychology **2**:*195. doi: 10.3389/fpsyg.2011.00195*

This article was submitted to Frontiers in Auditory Cognitive Neuroscience, a specialty of Frontiers in Psychology.

Copyright © 2011 Ott, Langer, Oechslin, Meyer and Jäncke. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.

3.2. Experiment 2: Processing of self-initiated speech sounds is different in musicians

Full Citation:

Ott, C. G. and Jäncke, L. (2013). Processing of self-initiated speech-sounds is different in musicians. *Frontiers in human neuroscience* 7 41.

3.2.1. Abstract

Musicians and musically untrained individuals have been shown to differ in a variety of functional brain processes such as auditory analysis and sensorimotor interaction. At the same time, internally operating forward models are assumed to enable the organism to discriminate the sensory outcomes of self-initiated actions from other sensory events by deriving predictions from efference copies of motor commands about forthcoming sensory consequences. As a consequence, sensory responses to stimuli that are triggered by a self-initiated motor act are suppressed relative to the same but externally-initiated stimuli, a phenomenon referred to as motor-induced suppression (MIS) of sensory cortical feedback. Moreover, MIS in the auditory domain has been shown to be modulated by the predictability of certain properties such as frequency or stimulus onset. The present study compares auditory processing of predictable and unpredictable self-initiated zero-delay speech sounds and piano tones between musicians and musical laymen by means of an event-related potential (ERP) and topographic pattern analysis (microstate analysis or EP mapping) approach. As in previous research on the topic of MIS, the amplitudes of the auditory event-related potential (AEP) N1 component were significantly attenuated for predictable and unpredictable speech sounds in both experimental groups to a comparable extent. On the other hand, AEP N1 amplitudes were enhanced for unpredictable self-initiated piano tones in both experimental groups similarly and MIS did not develop for predictable self-initiated piano tones at all. The more refined EP mapping revealed that the microstate exhibiting a typical auditory N1-like topography was significantly shorter in musicians when speech sounds and piano tones were self-initiated and predictable. In contrast, non-musicians only exhibited shorter auditory N1-like microstate durations in response to self-initiated and predictable piano tones. Taken together, our findings suggest that besides the known effect of MIS, internally operating forward models also

facilitate early acoustic analysis of complex tones by means of faster processing time as indicated by shorter auditory N1-like microstate durations in the first ~ 200 ms after stimulus onset. In addition, musicians seem to profit from this facilitation also during the analysis of speech sounds as indicated by comparable auditory N1-like microstate duration patterns between speech and piano conditions. In contrast, non-musicians did not show such an effect.

frontiers in
HUMAN NEUROSCIENCE

ORIGINAL RESEARCH ARTICLE
published: 22 February 2013
doi: 10.3389/fnhum.2013.00041

Processing of self-initiated speech-sounds is different in musicians

***Cyrill G. M. Ott*[1]* and Lutz Jäncke*[1,2]**

[1] Department of Psychology, Division Neuropsychology, Psychological Institute, University of Zurich, Zürich, Switzerland
[2] International Normal Aging and Plasticity Imaging Center, University of Zurich, Zürich, Switzerland

Edited by:
Eckart Altenmüller, University of Music and Drama Hannover, Germany

Reviewed by:
Kimmo Alho, University of Helsinki, Finland
Thomas F. Münte, University of Magdeburg, Germany
Peter Schneider, Heidelberg Medical School, Germany

***Correspondence:**
Cyrill G. M. Ott, Department of Psychology, Division Neuropsychology, Institute of Psychology, University Zurich, Binzmühlestrasse 14/Box 25, Zürich, CH-8050, Switzerland.
e-mail: c.ott@psychologie.uzh.ch

Musicians and musically untrained individuals have been shown to differ in a variety of functional brain processes such as auditory analysis and sensorimotor interaction. At the same time, internally operating forward models are assumed to enable the organism to discriminate the sensory outcomes of self-initiated actions from other sensory events by deriving predictions from efference copies of motor commands about forthcoming sensory consequences. As a consequence, sensory responses to stimuli that are triggered by a self-initiated motor act are suppressed relative to the same but externally initiated stimuli, a phenomenon referred to as motor-induced suppression (MIS) of sensory cortical feedback. Moreover, MIS in the auditory domain has been shown to be modulated by the predictability of certain properties such as frequency or stimulus onset. The present study compares auditory processing of predictable and unpredictable self-initiated 0-delay speech sounds and piano tones between musicians and musical laymen by means of an event-related potential (ERP) and topographic pattern analysis (TPA) [microstate analysis or evoked potential (EP) mapping] approach. As in previous research on the topic of MIS, the amplitudes of the auditory event-related potential (AEP) N1 component were significantly attenuated for predictable and unpredictable speech sounds in both experimental groups to a comparable extent. On the other hand, AEP N1 amplitudes were enhanced for unpredictable self-initiated piano tones in both experimental groups similarly and MIS did not develop for predictable self-initiated piano tones at all. The more refined EP mapping revealed that the microstate exhibiting a typical auditory N1-like topography was significantly shorter in musicians when speech sounds and piano tones were self-initiated and predictable. In contrast, non-musicians only exhibited shorter auditory N1-like microstate durations in response to self-initiated and predictable piano tones. Taken together, our findings suggest that besides the known effect of MIS, internally operating forward models also facilitate early acoustic analysis of complex tones by means of faster processing time as indicated by shorter auditory N1-like microstate durations in the first ~200 ms after stimulus onset. In addition, musicians seem to profit from this facilitation also during the analysis of speech sounds as indicated by comparable auditory N1-like microstate duration patterns between speech and piano conditions. In contrast, non-musicians did not show such an effect.

Keywords: motor-induced suppression, musicians, speech, plasticity, internal forward model

BACKGROUND

In the past 15 years, highly trained musicians have become an important model in neuroscience to study the effects of learning-induced cortical plasticity. It is now widely known and accepted that the intensive practice and training needed to achieve high musical skills leads to structural and functional short- and long-term alterations in the brain [for a review see Schlaug (2001); Münte et al. (2002); Jäncke (2009); Shahin (2011)]. There is also growing evidence that these alterations are not restricted particularly to structures and functions associated with perception and production of music, but also affect other domains such as language and speech (e.g., Patel, 2003; Thompson et al., 2003; Schön et al., 2004; Moreno and Besson, 2006; Besson et al., 2007; Marques et al., 2007; Parbery-Clark et al., 2009; Gordon et al., 2010; Schön et al., 2010; Besson et al., 2011; Colombo et al., 2011; Marie et al., 2011; Ott et al., 2011; Patel, 2011; Strait and Kraus, 2011; Kühnis et al., 2012).

However, in order to accomplish a high level of musical skill, the interplay between motor-actions and their resulting sensory consequences plays a critical role and thus comprises a substantial part of the daily training routines performed by adept musicians. On the other hand, various studies have shown that sensory responses to stimuli that are triggered by a self-initiated motor act are suppressed compared with the response to the same stimuli when they are externally triggered (e.g., Weiskrantz et al., 1971; Wolpert and Ghahramani, 2000; Martikainen et al.,

2005; Bäss et al., 2008; Aliu et al., 2009; Baess et al., 2009, 2011; Chen et al., 2012). To date, these functional differences are discussed to reflect the neuronal mechanisms providing the ability to discriminate sensory changes caused by one's own actions to those of others, thus enabling successful interaction with a highly complex sensory environment and efficient goal-directed behavior. Having this in mind, the question arises whether the high amount of sensorimotor practice and training that musicians are exposed to also leads to observable alterations with respect to the neural underpinnings of self- vs. externally generated sound discrimination.

Internal forward models of motor control (Wolpert and Kawato, 1998; Wolpert et al., 1998; Wolpert and Ghahramani, 2000; Wolpert and Flanagan, 2001) are addressing this differential processing and have integrated classical biological concepts such as the re-afference principle (Von Holst and Mittelstaedt, 1950; von Holst, 1954) or *corollary discharge* (Sperry, 1950). These models basically assume a prediction mechanism based on motor-to-sensory transformations within the central nervous system, in which an *efference copy* of the motor commands is translated into a representation of the expected sensory event (*corollary discharge*). This prediction encodes some of the expected sensory consequences of the movement to be executed and is then compared through sensory feedback loops to the sensory re-afference. The *efference copy* therefore enables the forward model to accurately predict the expected sensory feedback, resulting in a small prediction error, which in turn translates to a minimal response in the respective sensory cortex. In the absence of such an *efference copy* from the motor system, the forward model is unable to generate an accurate prediction of the sensory feedback, resulting in a larger prediction error and sensory field response, respectively. Moreover, the internal forward model mechanism is assumed to be a dynamic and adaptive modeling process based on estimations of the current state derived from the previous state (Kalveram, 1998, 2000). In other words, the sensory consequences based on the respective motor commands have to be learned by the forward model in order to generate an as accurate as possible prediction.

In human auditory cortex, this sensorimotor interplay can be readily observed. For example, the auditory cortex of a speaker responds to the sound of his own speech with a suppressed activation compared to the activation elicited by passive listening to mere playback of the speech, a phenomenon that has been labeled as *speaking-induced suppression* or SIS (Houde et al., 2002; Eliades and Wang, 2003). It also has been shown that a similar suppression phenomenon occurs in the somatosensory system, where responses to self-produced tactile stimuli are weaker relative to externally generated ones (Blakemore et al., 1998, 2000; Blakemore and Decety, 2001). These similarities suggest that the observed suppression effects reflect a more general property of sensory cortices in the sense that sensory feedback from any motor act is processed by comparing incoming feedback against a respective feedback-prediction derived from an *efference copy* of the motor command that produced the actual sensory feedback. This comparison in turn results in *motor-induced suppression* or MIS (Aliu et al., 2009).

With respect to suppression of auditory cortex activity, there is growing evidence from studies examining arbitrary pairings of a motor act with auditory stimuli. In 1973, Schafer and Marcus showed the human electroencephalographic (EEG) N1-response (peaking at ~100 ms) obtained at the vertex electrode to be significantly smaller to self-triggered auditory stimuli than to identical ones triggered by a machine. More recently, Martikainen et al. (2005) were able to demonstrate by using magnetoencephalography (MEG) that the human supratemporal cortex exhibits significant suppression when responding to self-initiated 1000 Hz sinusoid tones compared to passive listening to the same tones. Baess et al. (2011) used EEG and showed the N1-response to be even more attenuated to self- vs. externally triggered sounds when these were mixed within presentation blocks than the suppression observed in a traditional blocked design. Thus, the N1-suppression effect cannot be simply explained by contextual task differences and rather reflects the involvement of an internal prediction mechanism as stated above. Another study done by Bäss et al. (2008) revealed the N1-suppression effect to be attenuated but still present when sound onset and frequency are unpredictable. In particular, their findings show that forward model mechanisms tolerate uncertainties in the predictability of frequency and onset. In other words, forward models are readily operating as long as a subject is able to identify a sound as self-initiated. This in turn accounts mainly for interactions with unknown objects and/or insufficient visual information about the environment.

Anyhow, to our knowledge there is no study to date that addressed the earlier mentioned question whether intense musical training also leads to observable alterations in the neural underpinnings providing the discrimination of self- vs. externally generated sounds. To investigate this possibility, we compared high-density EEG-recordings of 16 professional pianists to those of 16 musically untrained individuals. We employed a paradigm similar to that used by Bäss et al. in their respective study of 2008, thus applying predictable and unpredictable outcomes of auditory consequences of a button press. Though, we introduced a speech and a piano condition and presented either CV-syllables ("Da" vs. "Ta"; speech condition) or piano tones ("C3" vs. "C5"; piano condition) instead of using pure sinusoid tones as employed in the afore-mentioned papers. The reason for using these stimuli was to elucidate whether possible musical training-induced differences in the magnitude of MIS are restricted to the sounds to which pianists are mainly exposed during practice or if those training effects are more widespread and are also affecting the domain of speech. To modulate the predictability of auditory feedback, the respective sounds were either fixed to the right vs. left response button (predictable outcome) or randomly assigned to the buttons with each button press (unpredictable outcome). By introducing an unpredictable condition, we wanted to verify a finding of one of our previous studies, in which we demonstrated that musicians process voiced and unvoiced stimuli similarly (Ott et al., 2011). According to this, predictability of voicing should affect the magnitude of prediction errors in the forward model of musicians to a lesser degree and thus not attenuate MIS of their auditory cortex to the same extent as of musical laymen.

Taking the most prominent findings of the respective literature into account, we first focused our analysis on amplitude modulations of the classical N1 auditory event-related potential (AEP)-component obtained at the vertex electrode (Schafer and Marcus, 1973; Numminen and Curio, 1999; Numminen et al., 1999; Curio et al., 2000; Houde et al., 2002; Heinks-Maldonado et al., 2005, 2006, 2007; Bäss et al., 2008; Aliu et al., 2009; Ventura et al., 2009; Baess et al., 2011). Here we hypothesized to find stronger suppression of the N1 amplitude in musicians in the predictable condition for speech and piano stimuli since their forward model should generate more precise predictions of the auditory consequences of a motor act due to their intense training, thus leading to smaller prediction errors resulting in stronger MIS. Based on the findings of Ott et al. (2011), we expected CV-syllable induced N1-amplitudes in the unpredictable condition to be comparable to the predictable condition in musicians, whereas N1-suppression should be attenuated in musical laymen when voicing of the syllables is unpredictable. On the other hand, unpredictable piano stimuli should induce attenuation of MIS to a similar extent in both groups.

In a second step, we will also analyze the spatial variations of the scalp voltage distribution across groups and conditions. This approach named Topographic Pattern Analysis (TPA) or evoked potential (EP) mapping has several advantages over conventional ERP analysis techniques. Using this technique we search for time segments of stable map topography that represent functional microstates of the brain. These microstates are assumed to reflect distinct information processing steps and provide particular advantages over the classical ERP analysis, such as experimenter and reference independence [for an overview, see e.g., Murray et al. (2008); Michel et al. (2009); Brunet et al. (2011)]. This kind of analysis exploits the high topographic resolution that high-density EEG-recordings provide. Using this approach, we searched for stable map topographies before, during, and after the N1 time window (similar as in our previous paper, Ott et al., 2011). Here, we are interested in examining whether the duration of these maps are different between musicians and non-musicians in the context of the suppression paradigm.

MATERIALS AND METHODS

SUBJECTS

Thirty-two healthy volunteers with normal audiological status and no history of neurological pathology were recruited for this study. Since the focus of this study was not on speech processing *per se* and for the sake of feasibility, we also accepted participants that were indeed german speaking but not native German or Swiss German speakers. However, we did not accept subjects of any tonal mother tongue, such as mandarin or other asian dialects. Sixteen pianists comprised the musician group (12 women, 4 men; mean age ± SD of 25.37 ± 4.44 years), with formal training starting at a mean age ± SD of 6.08 ± 0.94 years. All pianists were students, music teachers and/or members of an orchestra or band and practice piano playing on a daily basis of 1–7 h. The second group consisted of 16 musical laymen (13 women, 3 men; mean age ± SD of 25.12 ± 3.77 years) with no history of piano playing and no formal musical training exceeding the educational context of public elementary and secondary school. Regarding handedness, 4 participants within the pianist group were consistent left-handed, and 12 were consistent right-handed according to the Annett-Handedness-Questionnaire (Annett, 1970). The non-musician group comprised 2 consistent left-handed and 14 consistent right-handed subjects. To determine each participant's degree of musical competence, we applied the "Advanced Measures of Music Audiation" by Gordon (1989) prior to the EEG-experiment. Moreover, all participants also performed a short intelligence test (KAI) in order to rule out significant group differences in intelligence. **Table 1** summarizes the descriptive statistics of all criterion measures completed by the participants. All of them were paid for participation and gave informed written consent in accordance with procedures approved by the local ethics committee.

STIMULI

In this experiment, all participants either heard consonant-vowel (CV) syllables or piano tones. The CV-syllables consisted of a subset of speech stimuli already used in previous studies, such as Jäncke et al. (2002), Meyer et al. (2007), and Ott et al. (2011).

Table 1 | Descriptive statistics for the criterion measures completed by the subjects are listed group-wise (M, musicians; NM, non-musicians; GMA, Gordon Musical Aptitude; TS, total score; T, tonal score; R, rhythm score).

	Age		Gender		IQ		GMA		Handedness		
	Mean	SD	M	F	Mean	SD	Mean	SD	Left	Right	Ambi
M	25.37	4.44	4	12	124.6	6.5	TS*: 79.68 T*: 80.75 R*: 77.37	12.18 14.75 12.65	4	12	0
NM	25.12	3.77	3	13	128.4	11.9	TS*: 54.87 T*: 55.25 R*: 55.37	20.95 20.72 22.00	2	14	0

*Difference between experimental groups is significant at $p < 0.001$ [$T_{TS}(31) = 4.095$; $T_T(31) = 4.009$; $T_R(31) = 3.467$; note that T-Tests were only applied to variables "Age," "IQ," and "GMA." For "Gender" and "Handedness," Chi-Square tests were used]. All statistical tests shown in this table were conducted comparing the two experimental groups.

These syllables (/da/, /ta/) were digitally recorded by a trained phonetician at a sampling depth of 16-bit and a sampling rate of 44.1 kHz. Onset, duration, intensity, and fundamental frequency of the syllables were synchronized and edited by means of a speech editor. The articulatory release formed the criterion for temporal alignment of the syllables. Depending on their voice-onset-time (VOT), the duration of the stimuli ranged from 310 to 360 ms with a vowel duration of 300 ms (VOT's in ms for the stops were approximately "d" = 04 and "t" = 49). The piano stimuli comprised two piano tones of different pitch (/C3/, fundamental frequency at 130,813 Hz and /C5/, fundamental frequency at 523,251 Hz) generated with the Logic Express® software (Version 9.1.7, Apple® Inc; 95014 Cupertino, CA; http://www.apple.com/logicpro/). The duration of the tones was set to 360 ms. All auditory stimuli were presented binaurally at a sound pressure level of about 75 dB SPL using hi-fi headphones. Stimulus application and response recording was done using Presentation® software (Neurobehavioral Systems, USA).

EXPERIMENTAL SETUP

Based on previous research on self-initiated sounds (Schafer and Marcus, 1973; Martikainen et al., 2005; Bäss et al., 2008; Baess et al., 2011), we introduced three different tasks and a preceding training trial in our experiment:

In the motor-auditory (MA) task, subjects were instructed to generate an incidental key-press sequence by randomly pressing one or another out of two buttons on a computer keyboard with their left and right index fingers. Speech and piano stimuli were presented separately in three blocks at a time. Thus, each button press immediately triggered either a CV-syllable or a piano tone (0-delay stimulus application), respectively. The voiced CV-syllable (/da/) and the low-pitch piano tone (/C3/) were assigned to the left, whereas the unvoiced CV-syllable (/ta/) and the high-pitch piano tone (/C5/) were assigned to the right button. This assignment was not counterbalanced across subjects, as hearing a high piano tone after a left index finger and a low tone after a right index finger button press would be perceived as fairly disturbing, especially by pianists. To maintain consistency between speech and piano stimuli, these assignments were retained accordingly for both stimulus types. The subjects were asked to listen attentively to the sounds that were triggered by them. The interval between button presses was self-paced at a rate of about once every 3.5 s. Subjects also determined the sequential order of left vs. right button presses by themselves and were told to avoid estimating the intervals by internally counting seconds. Each button press was recorded to produce a trigger sequence for subsequent use in the auditory-only (A) task. Moreover, the Presentation® software counted the button presses in real-time and automatically ceased an ongoing block as soon as a minimum of 30 button presses per side (left vs. right index fingers) was reached. This resulted in a total of at least 90 self-initiated auditory stimulations with each CV-syllable and piano tone.

In the A task, the subjects listened to "externally" generated sequences of CV-syllables or piano tones, which used their own respective trigger sequences from the MA task. Using this approach, stimulation in the A task was identical to the MA task.

In the motor-only (M) task, subjects were again instructed to press the respective buttons in a self-paced interval and self-determined sequential order as in the MA task, but no sounds were delivered subsequently. This task served as a control condition to rule out the possibility of mere motor activity being responsible for any observed differences between the tasks involving auditory processing and between experimental groups as well. Hence, the M-task was subtracted from the other two tasks involving motor action prior to any further analysis.

To achieve unpredictability of the sensory consequence of a button press in the MA-unpredictable task (MAU), the respective assignments between buttons and auditory stimuli were randomly shuffled after each button press. Thus, the onset of a certain stimulus remained predictable (0-delay), whereas voicing of the CV-syllables (/da/ vs. /ta/) and pitch of the piano tones (/C3/ vs. /C5/) became unpredictable, respectively. Again, speech and piano stimuli were presented separately in three blocks at a time and the Presentation® software automatically ceased ongoing blocks after 30 self-paced/self-sequenced button presses per side.

Prior to the EEG-recordings, participants were trained to perform the self-paced rate of 3.5 s with visual feedback indicating whether their rate was too slow (>4.5 s), too fast (<2.5 s), or just right (2.5–4.5 s). During the experimental blocks, no visual feedback about their performance was given at all. As described, each task (MA, A, M, and MAU) comprised three blocks containing at least 60 button presses (30 left and 30 right) and/or auditory stimulations. All tasks were performed separately with speech and piano stimuli, resulting in a total amount of 1440 trials per subject. All blocks were presented in a randomized sequence over all participants. Though, as stimulation in the A task was dependent on the acoustical stimulation in the MA task, A task blocks could never precede their corresponding MA task blocks.

AEP RECORDINGS

Due to the fairly long duration of the experiment, we decided to split EEG-data acquisition in two separate recording sessions. EEG was recorded using a high-density Geodesic EEG system® (GSN300; Electrical Geodesic Inc., Oregon; http://www.egi.com) with 256-Channel HydroCel Geodesic Sensor Nets® (HCGSN120). Data was sampled at 500 Hz and band-pass filtered at 0.1–100 Hz. The vertex electrode (Cz) served as on-line recording reference and its exact location on each subject's scalp was noted down in order to ensure a preferably identical placement of the sensor net in the two recording sessions. Impedance was kept below 30 kOhm. During EEG recording, participants sat in a shielded, dimly lit room and a fixation cross was presented on an LCD-screen in front of them in order to reduce eye movements.

DATA ANALYSIS—EEG, N1 COMPONENT ANALYSIS

Each participant's EEG recordings were imported and analyzed in the BrainVision Analyzer2 software (Version 2.0.1; Brain Products GmbH, D-82205 Gilching; http://www.brainproducts.com). In a first step, data was band-pass filtered at 1–30 Hz. An ICA (independent component analysis) was then performed to

correct for ocular artifacts (e.g., Jung et al., 2000). Each EEG recording was visually inspected and trials with sweating and muscle artifacts, amplifier saturation, and remaining eye blinks or eye movements were excluded by means of a fully automatic raw data inspector. Noisy channels were interpolated and the data was then re-referenced to linked mastoid electrodes for ERP calculation. Each ERP waveform was an average of more than 60 repetitions of the EEG sweeps within a certain task evoked by the same auditory stimulus (MA, A, and MAU tasks) or button press (M task), respectively. EEG recordings were sectioned into 600 ms segments (200 ms pre-stimulus and 400 ms post-stimulus) and a baseline correction using the signal's pre-stimulus portion was carried out. To correct for mere motor activity in the MA and MAU tasks, difference waves between MA/MAU and M tasks were computed subsequently. Finally, ERP's for each stimulus and task were averaged for each subject and grand-averaged across subjects within the two groups separately.

In order to statistically confirm relevant differences between AEP's at Cz as a function of experimental group, task, and stimulus, mean amplitude event-related potentials (ERP's) time-locked to the auditory stimulation were measured in a specific latency window. This was individually pre-determined for each subject, task, and stimulus by visual inspection of the event-related signal. These individual latency windows were centered at the peak of the prominent N1 component and covered a total signal length of 10 ms around the center. Individually chosen latency windows were used to ensure that the mean amplitudes actually reflect the N1 peak values of every subject, task, and stimulus. Mean amplitudes were then averaged separately within groups (i.e., "pianists" vs. "non-musicians"), depending on the respective task (MA, A, MAU) and stimulus (/da/ vs. /ta/ and /C3/ vs. /C5/). Subsequently, a $2 \times 3 \times 2$ repeated measure Analysis of Variance (ANOVA) with a between-subject factor (*group*) and two within-subjects factors (*task* and *voiceness/pitch*) was computed for the central electrode (Cz). We used the multivariate approach to handle the problem of heteroscedasticity in repeated measurement designs (O'Brien and Kaiser, 1985). Thus, we will report F-values estimated from the multivariate Wilks' lambda statistic computed within the MANOVA. Subsequently, Bonferroni-Holm adjusted *post-hoc* t-tests were applied. The global level of significance was set at $p < 0.05$ for all statistical analyses.

DATA ANALYSIS—EEG, TOPOGRAPHIC PATTERN ANALYSIS (TPA)

TPA was performed for speech and piano stimuli separately. All subject- and grand-averaged ERP's were imported into the Cartool software (Version 3.51; The Cartool Community group, https://sites.google.com/site/cartoolcommunity/) and recalculated against the average reference. Then, the dominant map topographies on the scalp were identified by defining segments of stable voltage topography (or EP maps) using a topographic atomize and agglomerate hierarchical cluster analysis (T-AAHC) in the grand-averaged ERP's across tasks, stimuli, and groups over the full EEG segment length of 400 ms. These template maps are the mean maps over the period where the stable voltage topography segments were found. The clustering does not account for the latencies of maps, but only for their topographies and is done as a hypothesis generation step wherein the sequence of template maps that best accounts for the data is identified (Murray et al., 2008). The combination of a modified cross-validation and the Krzanowski–Lai criterion (e.g., Pascual-Marqui et al., 1995; Murray et al., 2008) was then used to determine the optimal number of templates. In accordance with the Cartool user guidelines (The Cartool Community group, https://sites.google.com/site/cartoolcommunity/user-s-guide), we identified the first on- and last offset of the typical N1-like map across tasks and groups by using the landscape obtained by the segmentation process. For speech stimuli, first on- and last offset of this map were at 0 and 192 ms, for piano stimuli they were at 0 and 202 ms. In a next step, we statistically verified the presence of each map that was found in the group-averaged data within that particular epoch over the same period in the ERP's of the individual subjects (i.e., "single-subject fitting"; Murray et al., 2008). This procedure is based on the spatial correlation between the template maps and the single-subject ERP data. This step allowed us to determine the duration of any given template map for each condition within the pianist and non-musician group between 0 and 192 ms (speech) and 0 and 202 ms (piano), respectively. These duration values were then statistically evaluated for each map of interest by means of a $2 \times 3 \times 2$ repeated measure ANOVA with the factors *group* (between-subjects), *task*, and *voicing/pitch* (within-subjects), as in the classical N1 component analysis. Subsequently, any significant interactions were further examined by applying Bonferroni–Holm adjusted *post-hoc* t-tests.

RESULTS

EEG DATA—N1 COMPONENT ANALYSIS (PEAK AMPLITUDES)—SPEECH

Grand-averaged waveforms evoked by CV-syllables recorded at Cz and the respective mean N1 peak amplitudes are illustrated in **Figure 1**. All stimuli elicited a prominent N1 component peaking at ~150 ms. Results of the $2 \times 3 \times 2$ repeated measures ANOVA revealed a significant main effect for the factor *task* [$F_{(1,\ 30)} = 6.974$, $p < 0.05$].

Bonferroni–Holm adjusted *post-hoc* t-tests indicated smaller amplitudes in the MA [$T_{(31)} = -3.433$, $p < 0.001$] and MAU [$T_{(31)} = -2.379$, $p < 0.05$] tasks compared to the A task. The difference between the MA and MAU tasks failed to reach significance [$T_{(31)} = 1.035$, $p > 0.05$]. Thus, a distinct attenuation to self-relative to externally initiated speech sounds was observed in general, replicating the results of various other studies addressing MIS and SIS, respectively (e.g., Schafer and Marcus, 1973; Houde et al., 2002; Eliades and Wang, 2003; Martikainen et al., 2005; Bäss et al., 2008; Baess et al., 2011).

Moreover, a significant main effect for the factor *voicing* was found in the ANOVA [$F_{(1,\ 30)} = 40.362$, $p < 0.001$], pointing to smaller amplitudes evoked by the unvoiced CV-syllable /ta/ than by the voiced CV-syllable /da/. Again, this finding is in line with the pre-existing literature covering the topic of sub-segmental speech processing (e.g., Simos et al., 1998; Sharma et al., 2000; Zaehle et al., 2007; Ott et al., 2011). However, neither significant interactions nor group-related effects were found with respect to N1 peak amplitudes elicited by speech.

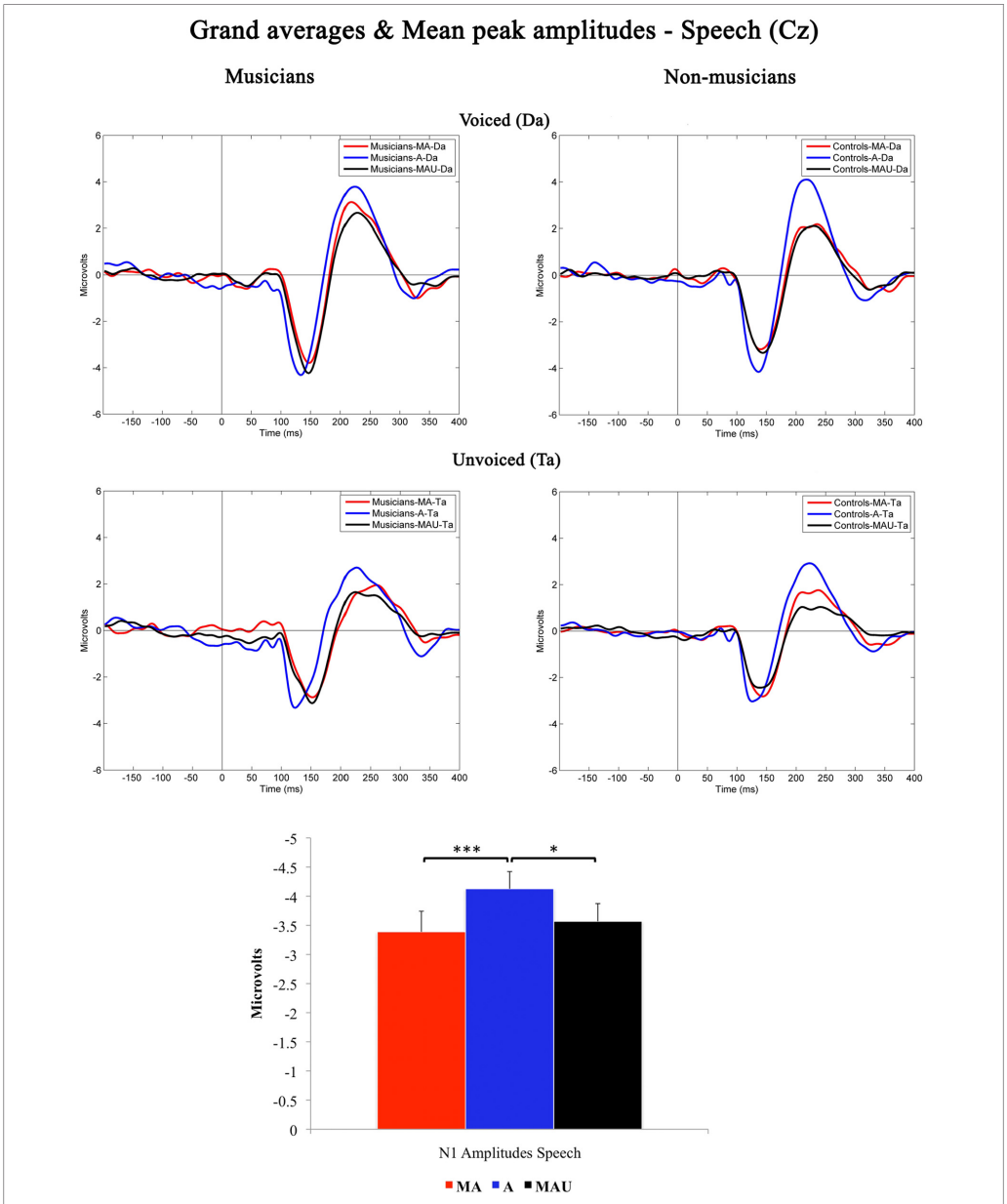

FIGURE 1 | Grand-averaged waveforms and corresponding mean N1 peak amplitudes evoked by voiced (top) and unvoiced (bottom) CV-syllables recorded at Cz for musicians (left) and non-musicians (right). MA, Motor-Auditory task; A, Auditory-Only task; MAU, Motor-Auditory unpredictable task. Mean N1 peak amplitudes are displayed irrespective of experimental group. *Difference is significant at $p < 0.05$, $T_{(31)} = -2.379$. ***Difference is significant at $p < 0.001$, $T_{(31)} = -3.433$. Error bars indicate standard errors. Time point "0" refers to the stimulus onset.

EEG DATA—N1 COMPONENT ANALYSIS (PEAK AMPLITUDES)—PIANO

Figure 2 shows the grand-averaged waveforms obtained at Cz elicited by low and high piano stimuli and the according mean N1 peak amplitudes. As found for CV-syllables, all piano stimuli evoked a prominent N1 component peaking at ~150 ms. The $2 \times 3 \times 2$ repeated measures ANOVA yielded significant main effects for the factors *task* [$F_{(1,\ 30)} = 4.408$, $p < 0.05$] and *pitch* [$F_{(1,\ 30)} = 2.129$, $p < 0.05$]. In addition, a significant interaction *task* by *pitch* [$F_{(1,\ 30)} = 4.426$, $p < 0.05$] was revealed. With respect to group-related differences regarding N1 peak amplitudes elicited by piano sounds, no significant effects were found.

Subsequent *post-hoc* analyses were applied for low and high piano tones separately, using Bonferroni–Holm adjusted *t*-tests. Low piano tones-evoked stronger N1 potentials in the MAU task relative to the A task [$T_{(31)} = 3.055$, $p < 0.01$], whereas high piano tones-evoked stronger potentials in the MAU task relative to the MA task [$T_{(31)} = 2.747$, $p < 0.01$]. All other comparisons failed to reach significance. Thus, no significant attenuation of mean N1 peak amplitudes to self-initiated piano sounds was observed.

EEG DATA—TOPOGRAPHIC PATTERN ANALYSIS—SPEECH

Figure 3 shows the speech-related results of the topographical EP mapping of the grand-averaged data for each group and task. Following the Cartool user guidelines (The Cartool Community group, https://sites.google.com/site/cartoolcommunity/user-s-guide), we first visually inspected the landscape and template maps resulting from the segmentation process to determine the particular time period and maps for the single-subject fitting. In order to provide consistency with the examination of the classical N1 component, we focused our analyses to the one template map that expressed a typical N1-like auditory topography (map 1) and therefore chose the time period for the single-subject fitting according to the first on- and last offset of this particular map. This resulted in a time window for the fitting procedure of 0–192 ms. Duration values of this map were then extracted from the fitted single-subject data in terms of timeframes and compared by means of a repeated measure ANOVA.

Results of the $2 \times 3 \times 2$ repeated measure ANOVA with between-subjects factor *group* and within-subjects factors *task* and *voicing* yielded a significant main effect for the factor *task* [$F_{(1,\ 30)} = 3.928$, $p < 0.05$] and a significant *task* by *group* interaction [$F_{(1,\ 30)} = 3.424$, $p < 0.05$].

Separate *post-hoc* analyses were conducted for the two experimental groups, using Bonferroni–Holm adjusted *t*-tests. Since no significant effect for the factor *voicing* was found, the respective auditory N1-like map (map 1) durations for the two CV-syllables (/da/ vs. /ta/) were averaged and analyzed together. *Post-hoc* *t*-tests revealed significantly longer durations for map 1 in the A task relative to the MA task in the pianist group [$T_{(15)} = 3.322$, $p < 0.01$]. Moreover, map 1 durations were longer in the MAU task compared to the MA task in the same group [$T_{(15)} = -2.214$, $p < 0.05$], though this particular difference failed to reach significance after correction for multiple comparisons. On the other hand, no significant differences between tasks regarding auditory N1-like map durations were found in the non-musician group at all.

EEG DATA—TOPOGRAPHIC PATTERN ANALYSIS—PIANO

Piano-tone related results of the topographical EP mapping are illustrated in **Figure 4**. The same procedure as with speech stimuli was used for the analysis. All analyses were again focused on the template map exhibiting a typical auditory N1-like topography (map 1) and the time window for the single-subject fitting thus ranged from 0 to 202 ms in accordance with first on- and last offsets of this particular map.

Here, the $2 \times 3 \times 2$ repeated measures ANOVA yielded significant main effects for the factors *task* [$F_{(1,\ 30)} = 3.832$, $p < 0.05$] and *group* [$F_{(1,\ 30)} = 4.572$, $p < 0.05$]. *Post-hoc* *t*-tests revealed significantly longer durations for map 1 in the MAU relative to the MA task [$T_{(31)} = -2.744$, $p < 0.01$] and in the A task relative to the MA task [$T_{(31)} = 1.719$, $p < 0.05$]. Though, the latter comparison failed to reach significance after correction for multiple comparisons. In addition, musicians exhibited longer auditory N1-like map durations in the MA and MAU tasks compared to non-musicians [$T_{(31)} = 2.099$, $p < 0.05$ and $T_{(31)} = 1.766$, $p < 0.05$, respectively]. Again, the latter comparison did not reach significance after correcting for multiple comparisons. Other than that, neither further main effects nor interactions were found.

DISCUSSION

The purpose of this study was to examine the extent to which intense musical training leads to functional alterations with respect to the neural underpinnings of the discrimination between self- vs. externally initiated sounds, a phenomenon known as *MIS* (e.g., Weiskrantz et al., 1971; Schafer and Marcus, 1973; Numminen and Curio, 1999; Numminen et al., 1999; Curio et al., 2000; Wolpert and Ghahramani, 2000; Houde et al., 2002; Heinks-Maldonado et al., 2005; Martikainen et al., 2005; Heinks-Maldonado et al., 2006, 2007; Ford et al., 2007; Bäss et al., 2008; Aliu et al., 2009; Baess et al., 2009, 2011; Chen et al., 2012). We were specifically interested in whether possible musical training-induced alterations were restricted to the sounds to which musicians are mainly exposed during their daily practice routines or if they are more widespread and are also affecting the processing of speech. Our hypothesis was that musical practice would have shaped the forward model of musicians in general such that it generates more precise predictions derived of motor-to-sensory transformations leading to smaller prediction errors and hence enhanced MIS of the auditory cortex. Based on the pre-existing literature addressing the topic of MIS in the auditory domain (e.g., Martikainen et al., 2005; Bäss et al., 2008; Aliu et al., 2009; Baess et al., 2009, 2011), we focused our analyses on the early auditory processing steps within the first 200 ms of stimulus onset and anticipated different neurophysiological activation patterns during these early processing stages.

In this study we used voiced and unvoiced CV-syllables (/da/ vs. /ta/, speech condition) and low- and high-pitched piano tones (/C3/ vs. /C5/, piano condition). Furthermore, we introduced a task in which the particular sensory consequence of a certain motor-action was unpredictable in terms of voicing

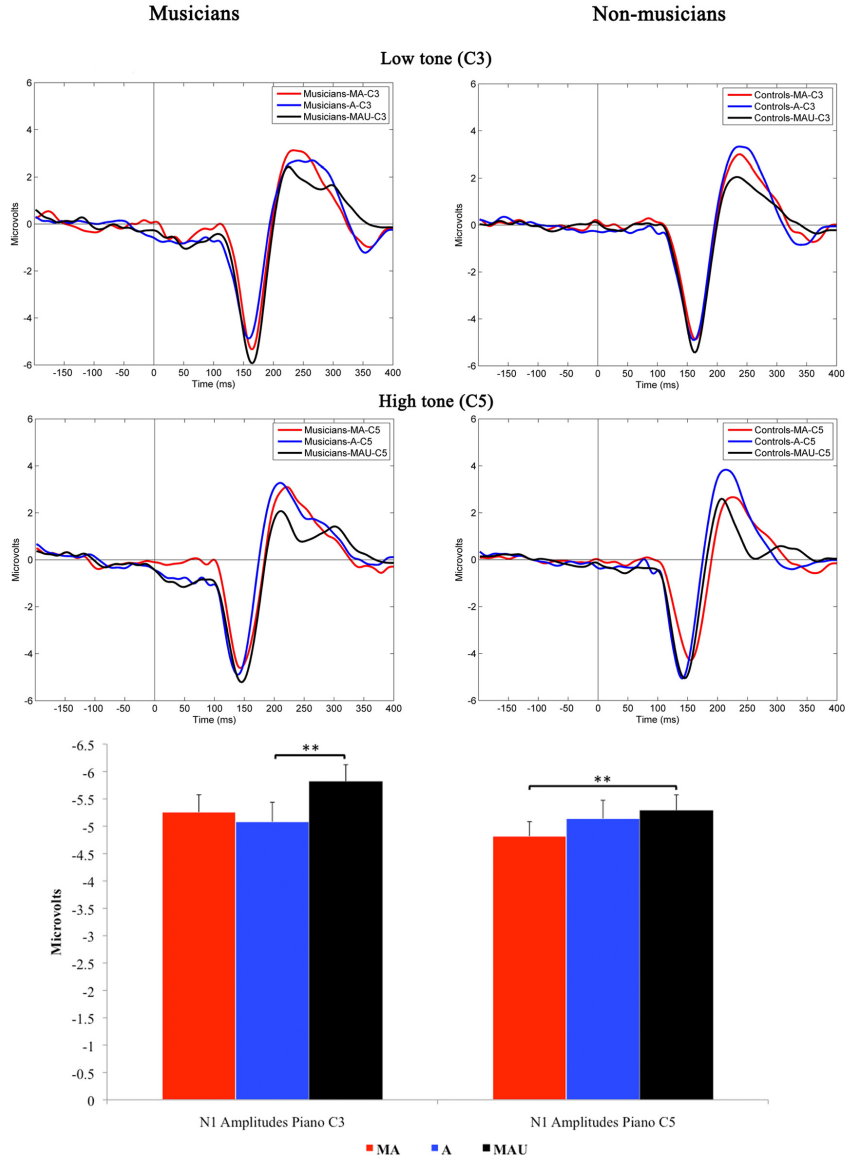

FIGURE 2 | Grand-averaged waveforms and corresponding mean N1 peak amplitudes evoked by low (top) and high (bottom) piano tones recorded at Cz for musicians (left) and non-musicians (right). MA, Motor-Auditory task; A, Auditory-Only task; MAU, Motor-Auditory unpredictable task. Mean N1 peak are displayed for low (left) and high (right) piano tones separately and irrespective of experimental group. **Differences are significant at $p < 0.01$, $T_{(31)} = 3.055$ and $p < 0.01$, $T_{(31)} = 2.747$, respectively. Error bars indicate standard errors. Time point "0" refers to the stimulus onset.

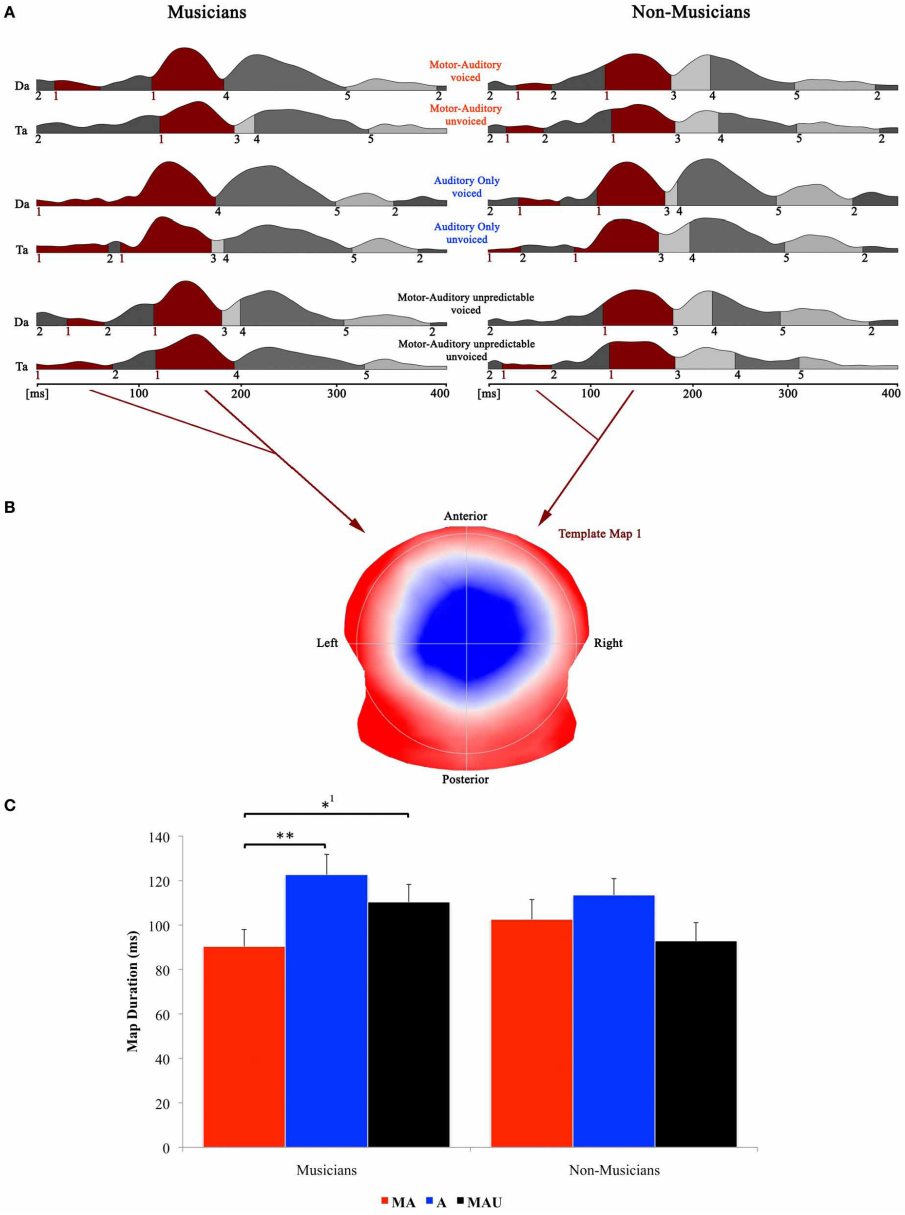

FIGURE 3 | Evoked potential data; speech condition. (A) Segments of stable map topography are shown group wise (left: musicians, right: non-musicians) for each condition under the global field power curve from 0 to 400 ms. The numbers below the segmentation landscape indicate the different template maps present in the grand-averaged data. Auditory N1-like template map 1 was found between 0 and 192 ms and was significantly shorter in the motor-auditory (MA) relative to the auditory-only (A) and

(Continued)

FIGURE 3 | Continued
motor-auditory unpredictable (MAU) tasks within the musician group. In the non-musician group, durations were comparable. **(B)** Map topography. Blue regions indicate negative, red regions indicate positive potential fields. **(C)** Duration of auditory N1-like template map (map 1) for musicians (left) and non-musicians (right) in the respective tasks. Durations are indicated irrespective of voicing. **Difference is significant at $p < 0.01$, $T_{(15)} = 3.322$. *[1] Difference is significant at $p < 0.05$, $T_{(15)} = -2.214$. Note that the difference indicated with *[1] failed to reach significance after correction for multiple comparisons. Also note that the data shown in panels (**A** and **B**) are derived of grand averaged waveforms, whereas map durations shown in panel **(C)** are obtained from individual ERPs. Scaling of template maps is normalized to values between −1 and 1 (for a detailed description of the microstate analysis procedure see e.g., Murray et al., 2008).

(speech condition) or pitch (piano condition). To control for mere motor activity possibly being responsible for any observed suppression effects involving auditory processing, a M task was also applied and corresponding difference waves between MA and M tasks were calculated for further analysis. High-density EEG recordings were derived during the early acoustic processing stages and analyzed in two different ways: Firstly we calculated and analyzed conventional AEP components and focused on the N1 component, which is particularly susceptible for MIS-effects (e.g., Schafer and Marcus, 1973; Numminen and Curio, 1999; Numminen et al., 1999; Curio et al., 2000; Houde et al., 2002; Heinks-Maldonado et al., 2005, 2006, 2007; Martikainen et al., 2005; Ford et al., 2007; Bäss et al., 2008; Aliu et al., 2009; Ventura et al., 2009; Baess et al., 2011). Complementary to the conventional ERP analysis we also applied a TPA. With this method we take advantage of the entire spatial information of our high-density EEG recording and recover a more detailed activation pattern. Using this approach, we focused on the microstate exhibiting a typical auditory N1-like topography and corresponding time windows of 0–192 ms (speech condition) and 0–202 ms (piano condition), respectively. As microstates and time windows are identified in an entirely data-driven approach substantially decreasing the subjective influence on the data analysis, this method can be considered a more objective method to analyze evoked electrical responses (Michel et al., 2009; Brunet et al., 2011).

SPEECH CONDITION

With respect to CV-syllables, we found that the N1 amplitude is significantly attenuated to self-initiated relative to externally initiated sounds in general, which is in line with the pre-existing literature covering MIS and SIS in particular (e.g., Houde et al., 2002; Eliades and Wang, 2003). Moreover, suppression of the N1 component is slightly reduced in the MAU relative to the MA task such as reported by Bäss et al. (2008), though the direct comparison between the two tasks failed to reach significance. In addition, durations of the microstate exhibiting a typical auditory N1-like topography are significantly shorter during early auditory processing stages (i.e., 0–192 ms) in the pianist group when CV-syllables are self-initiated and their outcome in terms of voicing is predictable compared to when they are externally initiated or voicing is unpredictable. On the other hand, there is no task-related difference between the respective microstate durations for non-musicians.

The findings with respect to suppression of the N1 amplitude in response to predictable and unpredictable self-initiated (speech) sounds mainly corresponds to the existing literature addressing this topic (e.g., Schafer and Marcus, 1973; Numminen and Curio, 1999; Numminen et al., 1999; Curio et al., 2000; Ford et al., 2001, 2007; Houde et al., 2002; Heinks-Maldonado et al., 2005, 2006, 2007; Martikainen et al., 2005; Bäss et al., 2008; Aliu et al., 2009; Ventura et al., 2009; Baess et al., 2011). Moreover, our results support the idea of an internal forward model deriving predictions about the sensory consequences of a certain motor-action (Wolpert and Kawato, 1998; Wolpert et al., 1998; Wolpert and Ghahramani, 2000; Wolpert and Flanagan, 2001) while tolerating uncertainties regarding the exact physical parameters of the expected sound (Bäss et al., 2008).

However, when comparing MIS of the N1 AEP component in pianists to the N1 suppression observed in musically untrained individuals, we could not confirm our hypothesis regarding the musicians internal forward model generating smaller prediction errors and thus enhanced MIS. As Aliu et al. (2009) described, MIS in the auditory cortex is most likely a learned phenomenon, that is, it is not immediately present but develops over time with repeated and consistent motor-action to sensory-consequence associations. Martikainen et al. (2005) clearly showed that MIS develops after only 60 successive motor-to-sensory couplings. Considering that all participants performed a training block consisting of 60 trials prior to the EEG-experiment, it could be the case that the forward model of musicians learns associations faster causing MIS to develop after fewer trials though the overall magnitude of N1 amplitude suppression obtained during the actual experimental blocks appears to be unaffected.

In fact, our findings concerning the TPA clearly demonstrate that the forward model of musicians actually alters the processing of a self-initiated motor-action's auditory consequences differently compared to the forward model of musical laymen, at least with respect to speech-sounds. The durations of the auditory N1-like template map are shorter in musicians during the first ∼200 ms when application of CV-syllables is self-initiated and their outcome in terms of voicing is predictable. This indicates that individuals exposed to intense musical training rely more on their internal forward model's predictions about basic acoustic features of auditory consequences of their own actions. Thus, the needed processing time of the actual auditory input during early stages is reduced, since no further processing is required after a positive match between the sensory goal and actual outcome. This would also explain why this facilitation effect diminishes when speech-sounds are self-initiated but their outcome in terms of voicing is unpredictable. In this case, the forward model's respective predictions are not as reliable causing the need for re-analysis of the actual auditory input which in turn leads to processing times comparable to the ones needed in conditions where speech-sounds are externally initiated. On the other hand,

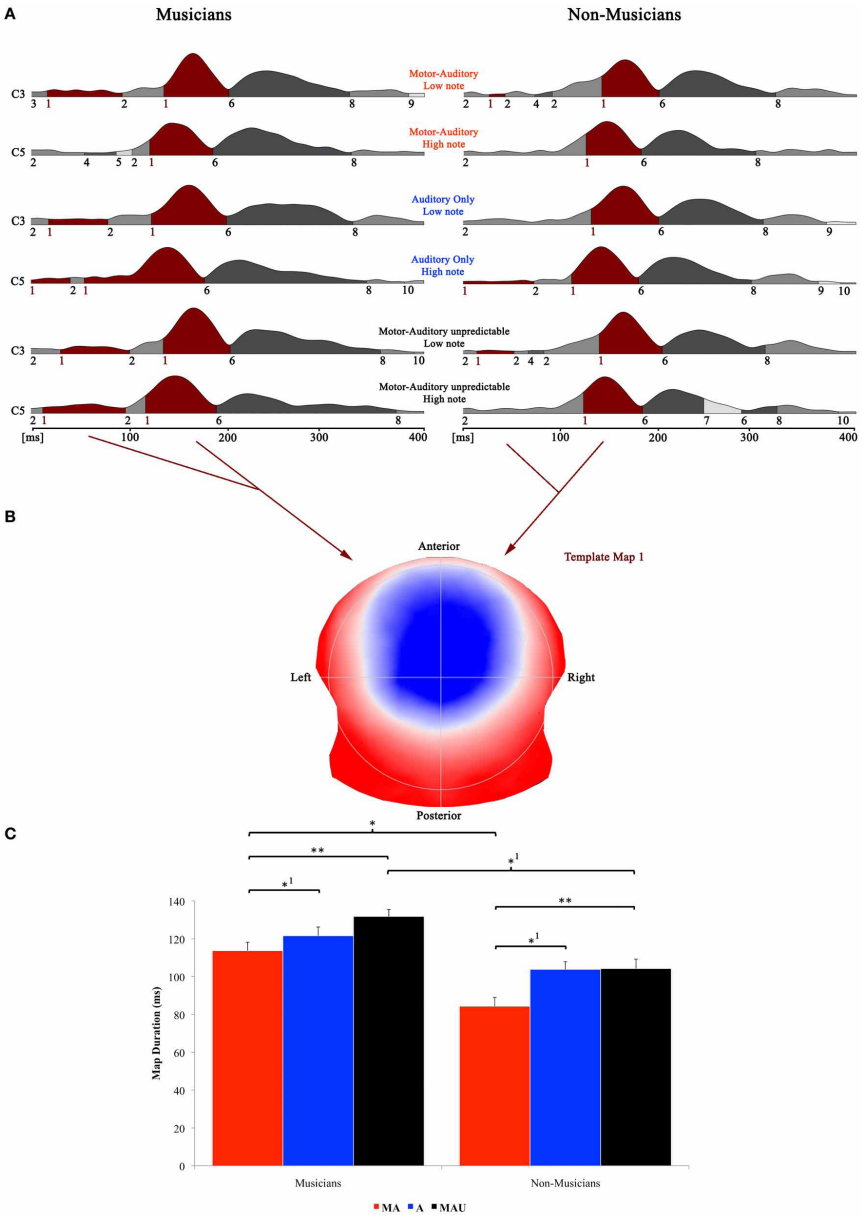

FIGURE 4 | Evoked potential data; piano condition. (A) Segments of stable map topography are shown group wise (left: musicians, right: non-musicians) for each condition under the global field power curve from 0 to 400 ms. The numbers below the segmentation landscape indicate the different template maps present in the grand-averaged data. Auditory N1-like template map 1 was found between 0 and 202 ms and was significantly longer in the motor-auditory unpredictable (MAU) relative to the

(Continued)

FIGURE 4 | Continued
motor-auditory (MA) task in both experimental groups. Map 1 durations were also longer in the auditory-only (A) relative to the MA task in both groups. In addition, musicians exhibited longer map 1 durations in the MA and MAU tasks compared to non-musicians. **(B)** Map topography. Blue regions indicate negative, red regions indicate positive potential fields. **(C)** Duration of auditory N1-like template map (map 1) for musicians (left) and non-musicians (right) in the respective tasks. Durations are indicated irrespective of pitch. **Difference is significant at, $p < 0.01$, $T_{(31)} = -2.744$. $^{+1}$Differences are significant at $p < 0.05$, $T_{(31)} = 1.719$ (MA vs. A) and $T_{(31)} = 1.766$ ($MAU_{musicians}$ vs. $MAU_{non-musicians}$). $^{+}$Difference is significant at $p < 0.05$, $T_{(31)} = 2.099$. Note that the differences indicated with $^{+1}$ failed to reach significance after correction for multiple comparisons. Also note that the data shown in panels (**A** and **B**) are derived of grand averaged waveforms, whereas map durations shown in panel **(C)** are obtained from individual ERPs. Scaling of template maps is normalized to values between −1 and 1 (for a detailed description of the microstate analysis procedure see e.g., Murray et al., 2008).

the forward model of musical laymen seems not to exhibit such a facilitation effect with respect to early auditory analysis of speech sounds, indicating that re-analysis is taking place even in case of a positive match.

Our second hypothesis, which was based on the results of our recent study about the processing of voiced and unvoiced acoustic stimuli in musicians (Ott et al., 2011), had also to be rejected. According to the findings of said study, musicians process voiced and unvoiced speech sounds similarly. Hence, we predicted that their forward model generates smaller prediction errors with voiced and unvoiced CV-syllables in the unpredictable task leading to a less prominent attenuation of MIS of the N1 AEP component compared to musical laymen. As our results show, this seems not to be the case. Moreover, we also failed to replicate the corresponding results concerning different physiological activation pattern between musicians and non-musicians in terms of AEP N1 amplitudes and auditory N1-like map durations. In particular, we found in our study of 2011 that N1 amplitudes and N1-like map durations in response to voiced and unvoiced auditory stimuli were comparable in musicians. In contrast, AEP N1 amplitudes were stronger and map durations were longer in non-musicians when processing voiced auditory stimuli relative to unvoiced ones. In the actual study however, both experimental groups showed the same activation pattern with respect to AEP N1 amplitudes and auditory N1-like map durations as the non-musician group in our 2011 study.

Given these discrepancies, the question arises to what extent the results of the two studies are actually comparable. Thus, it is instructive to point out the various differences between the two studies first and foremost with respect to experimental tasks and participants.

First, in our study of 2011 we applied a phonetic categorization task that required active processing of the acoustic properties discriminating voiced and unvoiced phonemes (e.g., VOT, formant-transitions, specific composition of formants, rises/falls in intensity, co-articulations). In opposition to this, the tasks used in our current study emphasize a rather passive/implicit processing of auditory speech stimuli. While the participants attention was mainly focused on maintaining the interval between button presses of ∼3 s, active/explicit processing of the auditory stimuli was not required to accomplish the task. At the same time, there are a variety of studies showing that auditory processing can be affected by specific task properties such as cognitive load, difficulty, repetition, and active vs. passive conditions (e.g., Lang and Kotchoubey, 2002; SanMiguel et al., 2008; Karns and Knight, 2009; Rao et al., 2010; Remijn and Kojima, 2010; Ben-David et al., 2011; Kam et al., 2011). From this perspective, it might be the case that the specific differences between musicians and non-musicians with respect to processing of voiced and unvoiced auditory stimuli only become apparent during conditions in which active/explicit processing of auditory input is required.

Second, the musician group in our study of 2011 comprised a high variability with respect to the genre and instruments the participating musicians mainly played whereas the musician group in our study at hand mostly consisted of classically educated pianists. As a matter of fact, it has been substantiated by various studies that musician's processing of sounds highly depends on instrument, genre, performance practice, and level of expertise (e.g., Koelsch et al., 1999; Pantev et al., 2001; Münte et al., 2003; Schneider et al., 2005; Vuust et al., 2005; Nikjeh et al., 2008; Vuust et al., 2009, 2012a,b). It is therefore quite conceivable that the processing of voicing in classic pianists is not even-handedly altered as in musicians of different genres playing different instruments and thus is rather comparable to the processing of voicing in non-musicians.

PIANO CONDITION

In contrast to speech stimuli, we found a different pattern concerning self- versus externally initiated piano tones. Here, N1 amplitudes are significantly enhanced in the MAU relative to the A and MA tasks in both experimental groups whereas no significant suppression occurs when piano sounds are self-initiated relative to when they are externally initiated. In particular, low-pitched piano tones elicit enhanced N1 amplitudes when self-triggered but unpredictable compared to externally triggered low-pitched piano tones (i.e., MAU > A). High-pitched piano tones on the other hand substantially enhance N1 amplitudes when self-triggered and unpredictable relative to self-triggered but predictable (i.e., MAU > MA) ones. Furthermore, we found that durations of the auditory N1-like template map are significantly longer during early auditory processing stages (i.e., 0–202 ms) in the pianist group in the MA task while in both groups the respective microstate durations are generally shorter when piano tones are self-triggered and predictable.

Our results at hand regarding self- vs. externally triggered piano tones seem somewhat contradictory to begin with, considering that most studies examining auditory processing of self- versus externally initiated sounds consistently showed motor-induced suppression of early AEP components (e.g., Schafer and Marcus, 1973; Numminen and Curio, 1999;

Numminen et al., 1999; Curio et al., 2000; Houde et al., 2002; Heinks-Maldonado et al., 2005, 2006, 2007; Martikainen et al., 2005; Ford et al., 2007; Bäss et al., 2008; Aliu et al., 2009; Ventura et al., 2009; Baess et al., 2011; Horváth et al., 2012). However, also contrasting findings were reported in several recent pitch-shifted ERP studies of self-produced vocalization (e.g., Behroozmand et al., 2009; Liu et al., 2010; Behroozmand and Larson, 2011; Chen et al., 2012). For example, Liu et al. (2010) compared neural responses to self-triggered pitch-shift stimuli (PSS) to those triggered by a computer during vocalization and listening. Their results showed that unpredictable self-initiated PSS elicited enhanced N1/P2 responses relative to unpredictable externally initiated PSS. Moreover, a similar study by Behroozmand and Larson (2011) reported N1 suppression effects to be strongly affected by the magnitude of PSS and being almost completely eliminated as the magnitude of PSS reached a certain level (i.e., 400 cents; 100 cents = 1 semitone in western music). In sum, these studies suggest that enhanced brain activity can be evoked to distinguish unpredictable self-triggered from unexpected externally triggered auditory stimulation. From this point of view, the general enhancement of the AEP N1 amplitude in response to unpredictable self-initiated piano tones at hand in our current study becomes less surprising. Moreover, larger prediction errors between efference copies and sensory re-afferences result in larger brain responses in the sensory cortex according to the forward model (Wolpert and Miall, 1996). Bearing in mind that the difference between the two piano tones that we used with respect to the fundamental frequency (F0) is fairly large (/C3/, F0 = 130,813 Hz and /C5/, F0 = 523,251 Hz), it might be the case that the enhancement of N1 amplitudes in response to unpredictable piano tones that we observed in the current study is simply due to large mismatches between the forward model's predictions and the actual sensory feedback.

Anyhow, this still does not explain why AEP N1 amplitudes in response to predictable self-initiated piano tones do not exhibit significant suppression relative to predictable externally initiated ones in our current study. Though, a plausible explanation can be derived from the results of a recent MEG study: Aliu et al. (2009) reported suppressed N1m amplitudes in response to short (100 ms), binaurally presented simple 0-delay 1 kHz-tones. In particular, MIS development was clearly present in the left, but did not extend to the right hemisphere. According to the asymmetric sampling in time model ("AST-model"; Poeppel, 2003; Hickok and Poeppel, 2007; Poeppel et al., 2008; Luo and Poeppel, 2012), temporal analysis and integration of "long-scale" auditory input signals (i.e., ~150–300 ms) mainly carrying spectral information preferentially drive the right hemisphere. Hence, it is fairly conceivable that MIS for predictable self-initiated piano tones did not develop to an extent where it can induce observable and statistically relevant differences with respect to AEP N1 amplitudes.

Despite the absence of significant suppression effects of AEP N1 amplitudes, the results of the corresponding TPA reveal alterations of auditory processing of self-initiated piano tones in both experimental groups. Auditory N1-like map durations are significantly shorter in pianists as well as in musical laymen when the piano tones are self-initiated and predictable. Again, this indicates a facilitation effect of accurate sensory predictions by the forward model such as less processing time is needed for early auditory analysis. Moreover, pianists show enhanced N1-like map durations relative to non-musicians in the MA task, suggesting a more refined auditory analysis and increased neuronal representation of self-initiated and predictable piano tones in pianists. This finding fits in well with the results of various other studies showing enhanced auditory processing of complex tones in musicians (e.g., Pantev et al., 1998; Shahin et al., 2005, 2007; Kuriki et al., 2006; Baumann et al., 2008).

However, it is also of importance to briefly address one additional issue at this point: the auditory N1 has been shown to vary as a function of attention. Thus, it might be the case that possible differences in MIS between the groups are corrupted by mere attention effects. Though, the question whether attention influences the N1-suppression effect for self-initiated sounds has recently been addressed by Timm et al. (2013). Their results showed the N1 itself to be affected by attention, but there was no interaction between attention and self-initiation effects. The authors therefore conclude that the N1-suppression effect for self-initiated sounds is independent of attention. Hence, we assume that attention does not account for the effects we observed in the present study.

CONCLUSION

Our findings indicate that intense musical training facilitates the processing of predictable self-initiated speech sounds in terms of faster early auditory analysis in the first ~200 ms. In particular, pianists show a comparable auditory N1-like template map duration pattern in response to predictable self-initated speech and piano tones, whereas a substantial reduction of processing time in non-musicians is only present in response to predictable self-initiated piano tones. Moreover, pianists and non-musicians did not significantly differ with respect to suppression of AEP N1 amplitudes neither in response to self-initiated speech sounds nor piano tones. Taken together, our results suggest that pianists rely on their forward model's predictions of sensory outcomes to the same extent when processing speech sounds and piano tones. In contrast, non-musicians seem to analyze predictable self-initiated speech sounds in a more elaborate manner even in the case of a positive match between their forward model's predictions and the actual auditory input. In addition, internal forward mechanisms do not mandatorily lead to suppressed cortical feedback to self-initiated complex sounds but can take beneficial effect with respect to processing time, as indicated by the absence of significant suppression of AEP N1 amplitudes and nonetheless shortened auditory N1-like map durations in response to predictable self-initiated piano tones.

ACKNOWLEDGMENTS

This study was supported by a grant of the Swiss National Science Foundation (SNF) to Lutz Jäncke (32003B_121927).

REFERENCES

Aliu, S. O., Houde, J. F., and Nagarajan, S. S. (2009). Motor-induced suppression of the auditory cortex. *J. Cogn. Neurosci.* 21, 791–802.

Annett, M. (1970). A classification of hand preference by association analysis. *Br. J. Psychol.* 61, 303–321.

Baess, P., Horváth, J., Jacobsen, T., and Schröger, E. (2011). Selective suppression of self-initiated sounds in an auditory stream: an ERP study. *Psychophysiology* 48, 1276–1283.

Baess, P., Widmann, A., Roye, A., Schröger, E., and Jacobsen, T. (2009). Attenuated human auditory middle latency response and evoked 40-Hz response to self-initiated sounds. *Eur. J. Neurosci.* 29, 1514–1521.

Bäss, P., Jacobsen, T., and Schröger, E. (2008). Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: evidence for internal forward models with dynamic stimulation. *Int. J. Psychophysiol.* 70, 137–143.

Baumann, S., Meyer, M., and Jäncke, L. (2008). Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *J. Cogn. Neurosci.* 20, 2238–2249.

Behroozmand, R., Karvelis, L., Liu, H., and Larson, C. R. (2009). Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation. *Clin. Neurophysiol.* 120, 1303–1312.

Behroozmand, R., and Larson, C. R. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neuroscience* 12:54. doi: 10.1186/1471-2202-12-54

Ben-David, B., Campeanu, S., Tremblay, K., and Alain, C. (2011). Auditory evoked potentials dissociate rapid perceptual learning from task repetition without learning. *Psychophysiology* 48, 797–807.

Besson, M., Chobert, J., and Marie, C. (2011). Transfer of training between music and speech: common processing, attention, and memory. *Front. Psychol.* 2:94. doi: 10.3389/fpsyg.2011.00094

Besson, M., Schön, D., Moreno, S., Santos, A., and Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restor. Neurol. Neurosci.* 25, 399–410.

Blakemore, S. J., and Decety, J. (2001). From the perception of action to the understanding of intention. *Nat. Rev. Neurosci.* 2, 561–567.

Blakemore, S. J., Wolpert, D., and Frith, C. (2000). Why can't you tickle yourself? *Neuroreport* 11, R11–R16.

Blakemore, S. J., Wolpert, D. M., and Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nat. Neurosci.* 1, 635–640.

Brunet, D., Murray, M. M., and Michel, C. M. (2011). Spatiotemporal analysis of multichannel EEG: cartool. *Comput. Intell. Neurosci.* 2011:813870. doi: 10.1155/2011/813870

Chen, Z., Chen, X., Liu, P., Huang, D., and Liu, H. (2012). Effect of temporal predictability on the neural processing of self-triggered auditory stimulation during vocalization. *BMC Neuroscience* 13:55. doi: 10.1186/1471-2202-13-55

Colombo, L., Deguchi, C., Boureux, M., Sarlo, M., and Besson, M. (2011). Detection of pitch violations depends upon the familiarity of intonational contour of sentences. *Cortex* 47, 557–568.

Curio, G., Neuloh, G., Numminen, J., Jousmäki, V., and Hari, R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Hum. Brain Mapp.* 9, 183–191.

Eliades, S. J., and Wang, X. (2003). Sensory-motor interaction in the primate auditory cortex during self-initiated vocalizations. *J. Neurophysiol.* 89, 2194–2207.

Ford, J. M., Mathalon, D. H., Heinks, T., Kalba, S., Faustman, W. O., and Roth, W. T. (2001). Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. *Am. J. Psychiatry* 158, 2069–2071.

Ford, J. M., Gray, M., Faustman, W. O., Roach, B. J., and Mathalon, D. H. (2007). Dissecting corollary discharge dysfunction in schizophrenia. *Psychophysiology* 44, 522–529.

Gordon, E. E. (1989). *Advanced Measures of Music Audiation*. Chicago, IL: GIA Publications, Inc.

Gordon, R. L., Schön, D., Magne, C., Astésano, C., and Besson, M. (2010). Words and melody are intertwined in perception of sung words: EEG and behavioral evidence. *PLoS ONE* 5:e9889. doi: 10.1371/journal.pone.0009889

Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., and Ford, J. M. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology* 42, 180–190.

Heinks-Maldonado, T. H., Mathalon, D. H., Houde, J. F., Gray, M., Faustman, W. O., and Ford, J. M. (2007). Relationship of imprecise corollary discharge in schizophrenia to auditory hallucinations. *Arch. Gen. Psychiatry* 64, 286–296.

Heinks-Maldonado, T. H., Nagarajan, S. S., and Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport* 17, 1375–1379.

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.

Horváth, J., Maess, B., Baess, P., and Tóth, A. (2012). Action-sound coincidences suppress evoked responses of the human auditory cortex in EEG and MEG. *J. Cogn. Neurosci.* 24, 1919–1931.

Houde, J. F., Nagarajan, S. S., Sekihara, K., and Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: an MEG study. *J. Cogn. Neurosci.* 14, 1125–1138.

Jäncke, L. (2009). Music drives brain plasticity. *F1000 Biol. Rep.* 1:78. doi: 10.3410/B1-78

Jäncke, L., Wüstenberg, T., Scheich, H., and Heinze, H. J. (2002). Phonetic perception and the temporal cortex. *Neuroimage* 15, 733–746.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* 37, 163–178.

Kalveram, K. T. (1998). *Wie das Individuum mit Seiner Umwelt Interagiert. Psychologische, Biologische und Kybernetische Betrachtungen über die Funktion von Verhalten, Chapters 1 and 4*. Lengerich: Pabst, Science Publishers.

Kalveram, K. T. (2000). "Sensorimotor sequential learning by a neural network based on redefined Hebbian Learning," in *Artificial Neural Networks in Medicine and Biology*, eds H. Malmgren, M. Borga, and L. Niklasson (London: Springer), 271–276.

Kam, J. W., Dao, E., Farley, J., Fitzpatrick, K., Smallwood, J., Schooler, J. W., et al. (2011). Slow fluctuations in attentional control of sensory cortex. *J. Cogn. Neurosci.* 23, 460–470.

Karns, C. M., and Knight, R. T. (2009). Intermodal auditory, visual, and tactile attention modulates early stages of neural processing. *J. Cogn. Neurosci.* 21, 669–683.

Koelsch, S., Schröger, E., and Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *Neuroreport* 10, 1309–1313.

Kühnis, J., Elmer, S., Meyer, M., and Jäncke, L. (2012). Musicianship boosts perceptual learning of pseudoword-chimeras: an electrophysiological approach. *Brain Topogr.* 26, 110–125.

Kuriki, S., Kanda, S., and Hirata, Y. (2006). Effects of musical experience on different components of MEG responses elicited by sequential piano-tones and chords. *J. Neurosci.* 26, 4046–4053.

Lang, S., and Kotchoubey, B. (2002). Brain responses to number sequences with and without active task requirement. *Clin. Neurophysiol.* 113, 1734–1741.

Liu, H., Behroozmand, R., and Larson, C. R. (2010). Enhanced neural responses to self-triggered voice pitch feedback perturbations. *Neuroreport* 21, 527–531.

Luo, H., and Poeppel, D. (2012). Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. *Front. Psychol.* 3:170. doi: 10.3389/fpsyg.2012.00170

Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., and Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in mandarin chinese. *J. Cogn. Neurosci.* 23, 2701–2715.

Marques, C., Moreno, S., Castro, S. L., and Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. *J. Cogn. Neurosci.* 19, 1453–1463.

Martikainen, M. H., Kaneko, K., and Hari, R. (2005). Suppressed responses to self-triggered sounds in the human auditory cortex. *Cereb. Cortex* 15, 299–302.

Meyer, M., Elmer, S., Baumann, S., and Jancke, L. (2007). Short-term plasticity in the auditory system: differential neural responses to perception and imagery of speech and music. *Restor. Neurol. Neurosci.* 25, 411–431.

Michel, C. M., Koenig, Th., Brandeis, D., Gianotti, L. R. R., and Wackermann, J. (2009). "Electrical neuroimaging in the time domain - microstate analysis," in *Electrical Neuroimaging. Chapter 6*, eds C. M. Michel, Th. Koenig, D. Brandeis, L. R. R. Gianotti, and J. Wackermann (Cambridge: Cambridge University Press), 123–125.

Moreno, S., and Besson, M. (2006). Musical training and language-related brain electrical activity in children. *Psychophysiology* 43, 287–291.

Münte, T. F., Altenmüller, E., and Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nat. Rev. Neurosci.* 3, 473–478.

Münte, T. F., Nager, W., Beiss, T., Schroeder, C., and Altenmüller, E. (2003). Specialization of the specialized: electrophysiological investigations in professional musicians. *Ann. N.Y. Acad. Sci.* 999, 131–139.

Murray, M. M., Brunet, D., and Michel, C. M. (2008). Topographic ERP analyses: a step-by-step tutorial review. *Brain Topogr.* 20, 249–264.

Nikjeh, D. A., Lister, J. J., and Frisch, S. A. (2008). Hearing of note: an electrophysiologic and psychoacoustic comparison of pitch discrimination between vocal and instrumental musicians. *Psychophysiology* 45, 994–1007.

Numminen, J., and Curio, G. (1999). Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neurosci. Lett.* 272, 29–32.

Numminen, J., Salmelin, R., and Hari, R. (1999). Subject's own speech reduces reactivity of the human auditory cortex. *Neurosci. Lett.* 265, 119–122.

O'Brien, R. G., and Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: an extensive primer. *Psychol. Bull.* 97, 316–333.

Ott, C. G. M., Langer, N., Oechslin, M. S., and Meyer, M. (2011). Processing of voiced and unvoiced acoustic stimuli in musicians. *Front. Psychology* 2:195. doi: 10.3389/fpsyg.2011.00195

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., and Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature* 392, 811–814.

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., and Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport* 12, 169–174.

Parbery-Clark, A., Skoe, E., Lam, C., and Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear Hear.* 30, 653–661.

Pascual-Marqui, R. D., Michel, C. M., and Lehmann, D. (1995). Segmentation of brain electrical activity into microstates: model estimation and validation. *IEEE Trans. Biomed. Eng.* 42, 658–665.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nat. Neurosci.* 6, 674–681.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The opera hypothesis. *Front. Psychology* 2:142. doi: 10.3389/fpsyg.2011.00142

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time.' *Speech Commun.* 41, 245–255.

Poeppel, D., Idsardi, W. J., and van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 1071–1086.

Rao, A., Zhang, Y., and Miller, S. (2010). Selective listening of concurrent auditory stimuli: an event-related potential study. *Hear. Res.* 268, 123–132.

Remijn, G. B., and Kojima, H. (2010). Active versus passive listening to auditory streaming stimuli: a near-infrared spectroscopy study. *J. Biomed. Opt.* 15, 037006.

SanMiguel, I., Corral, M. J., and Escera, C. (2008). When loading working memory reduces distraction: behavioral and electrophysiological evidence from an auditory-visual distraction paradigm. *J. Cogn. Neurosci.* 20, 1131–1145.

Schafer, E. W., and Marcus, M. M. (1973). Self-stimulation alters human sensory brain responses. *Science* 181, 175–177.

Schlaug, G. (2001). The brain of musicians. A model for functional and structural adaptation. *Ann. N.Y. Acad. Sci.* 930, 281–299.

Schneider, P., Sluming, V., Roberts, N., Bleeck, S., and Rupp, A. (2005). Structural, functional, and perceptual differences in Heschl's gyrus and musical instrument preference. *Ann. N.Y. Acad. Sci.* 1060, 387–394.

Schön, D., Gordon, R., Campagne, A., Magne, C., Astésano, C., Anton, J., et al. (2010). Similar cerebral networks in language, music and song perception. *Neuroimage* 51, 450–461.

Schön, D., Magne, C., and Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* 41, 341–349.

Shahin, A., Roberts, L. E., Pantev, C., Trainor, L. J., and Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport* 16, 1781–1785.

Shahin, A. J. (2011). Neurophysiological influence of musical training on speech perception. *Front. Psychology* 2:126. doi: 10.3389/fpsyg.2011.00126

Shahin, A. J., Roberts, L. E., Pantev, C., Aziz, M., and Picton, T. W. (2007). Enhanced anterior-temporal processing for complex tones in musicians. *Clin. Neurophysiol.* 118, 209–220.

Sharma, A., Marsh, C. M., and Dorman, M. F. (2000). Relationship between N1 evoked potential morphology and the perception of voicing. *J. Acoust. Soc. Am.* 108, 3030–3035.

Simos, P. G., Diehl, R. L., Breier, J. I., Molis, M. R., Zouridakis, G., and Papanicolaou, A. C. (1998). MEG correlates of categorical perception of a voice onset time continuum in humans. *Cogn. Brain Res.* 7, 215–219.

Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *J. Comp. Physiol. Psychol.* 43, 482–489.

Strait, D., and Kraus, N. (2011). Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Front. Psychology* 2:113. doi: 10.3389/fpsyg.2011.00113

Thompson, W. F., Schellenberg, E. G., and Husain, G. (2003). Perceiving prosody in speech. Effects of music lessons. *Ann. N.Y. Acad. Sci.* 999, 530–532.

Timm, J., SanMiguel, I., Saupe, K., and Schröger, E. (2013). The N1-suppression effect for self-initiated sounds is independent of attention. *BMC Neuroscience* 14:2. doi: 10.1186/1471-2202-14-2

Ventura, M. I., Nagarajan, S. S., and Houde, J. F. (2009). Speech target modulates speaking induced suppression in auditory cortex. *BMC Neuroscience* 10:58. doi: 10.1186/1471-2202-10-58

von Holst, E. (1954). Relations between the central nervous system and the peripheral organs. *Br. J. Anim. Behav.* 2, 89–94.

Von Holst, E., and Mittelstaedt, H. (1950). Das Reafferenzprinzip [The Reafference Principle]. *Naturwissenschaften* 37, 464–476.

Vuust, P., Brattico, E., Seppänen, M., Näätänen, R., and Tervaniemi, M. (2012a). Practiced musical style shapes auditory skills. *Ann. N.Y. Acad. Sci.* 1252, 139–146.

Vuust, P., Brattico, E., Seppänen, M., Näätänen, R., and Tervaniemi, M. (2012b). The sound of music: differentiating musicians using a fast, musical multi-feature mismatch negativity paradigm. *Neuropsychologia* 50, 1432–1443.

Vuust, P., Ostergaard, L., Pallesen, K. J., Bailey, C., and Roepstorff, A. (2009). Predictive coding of music–brain responses to rhythmic incongruity. *Cortex* 45, 80–92.

Vuust, P., Pallesen, K. J., Bailey, C., van Zuijen, T. L., Gjedde, A., Roepstorff, A., et al. (2005). To musicians, the message is in the meter pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *Neuroimage* 24, 560–564.

Weiskrantz, L., Elliott, J., and Darlington, C. (1971). Preliminary observations on tickling oneself. *Nat. Neurosci.* 230, 598–599.

Wolpert, D. M., and Flanagan, J. R. (2001). Motor prediction. *Curr. Biol.* 11, R729–R732.

Wolpert, D. M., and Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nat. Neurosci.* 3(Suppl.), 1212–1217.

Wolpert, D. M., and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Netw.* 11, 1317–1329.

Wolpert, D. M., and Miall, R. C. (1996). Forward models for physiological motor control. *Neural Netw.* 9, 1265–1279.

Wolpert, D. M., Miall, R. C., and Kawato, M. (1998). Internal models in the cerebellum. *Trends Cogn. Sci.* 2, 338–347.

Zaehle, T., Jancke, L., and Meyer, M. (2007). Electrical brain imaging evidences left auditory cortex involvement in speech and non-speech discrimination based on temporal features. *Behav. Brain Funct.* 3, 63.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 27 November 2012; paper pending published: 18 December 2012; accepted: 04 February 2013; published online: 22 February 2013.

Citation: Ott CGM and Jäncke L. (2013) Processing of self-initiated speech-sounds is different in musicians. Front. Hum. Neurosci. 7:41. doi: 10.3389/fnhum. 2013.00041

Copyright © 2013 Ott and Jäncke. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.

3.3. Experiment 3: Musical expertise affects attention as reflected by auditory evoked gamma-band activity in human EEG

Accepted for Publication in *Neuroreport* (26 Feb. 2013)

3.3.1. Abstract

Musical expertise has been shown to induce widespread structural and functional alterations in the brain, even-handedly affecting top-down and bottom-up factors. At the same time, it is known that the early evoked gamma-band response (GBR) can be modulated by top-down as well as bottom-up factors such as attention and sound intensity. In this study, we investigated the effects of musicianship and attention on the intensity modulation of the auditory evoked GBR.

We compared the electroencephalogram (EEG) of 17 professional musicians to that of 17 musical laymen obtained during either a forced-choice discrimination task (active) or a passive listening condition. Pure 1000 Hz sine tones were presented at three systematically varied sound intensities (40, 60 and 80 dB SPL).

The results concerning auditory evoked potentials (AEP's) and evoked GBR's obtained in the active condition predominantly corresponded to the findings of previous studies. Beside the already known augmentation of the early evoked GBR due to enhanced inter-trial phase coherence with increasing sound intensity, we also observed stronger GBR's and enhanced phase-locking in the active condition compared to passive listening, while the general shape of intensity modulation was comparable between the two conditions. In addition, phase-locking to stimulus onset was increased for stimuli of all three intensities when attended, whereas in musicians only stimuli of the highest intensity (80 dB) induced significantly increased phase-locking in the active condition.

Taken together, our results suggest that musical expertise influences attention effects on the intensity-modulated early auditory evoked GBR with respect to inter-trial phase-coherence.

Musical expertise affects attention as reflected by auditory evoked gamma-band activity in human EEG

Cyrill G. M. Ott[1], Christina Stier[1], Christoph S. Herrmann[3,4] and Lutz Jäncke[1,2]*

[1]Division Neuropsychology, Psychological Institute, University of Zurich, Switzerland

[2]International Normal Aging and Plasticity Imaging Center (INAPIC), University of Zurich,
Switzerland

[3]Experimental Psychology Lab, Institute of Psychology, Carl von Ossietzky University, Oldenburg, Germany

[4]Research Center Neurosensory Science, Carl von Ossietzky University, Oldenburg, Germany

*Address for correspondence
Cyrill G. M. Ott, lic. phil.
Division Neuropsychology
Institute of Psychology
University Zurich
Binzmühlestrasse 14 / Box 25
CH-8050 Zürich
Switzerland
Tel.: +41 (44) 635 74 07
Fax: +41 (44) 635 74 09
Email address:
c.ott@psychologie.uzh.ch

Abstract

Musical expertise has been shown to induce widespread structural and functional alterations in the brain, even-handedly affecting top-down and bottom-up factors. At the same time, it is known that the early evoked gamma-band response (GBR) can be modulated by top-down as well as bottom-up factors such as attention and sound intensity. In this study, we investigated the effects of musicianship and attention on the intensity modulation of the auditory evoked GBR.

We compared the electroencephalogram (EEG) of 17 professional musicians to that of 17 musical laymen obtained during either a forced-choice discrimination task (active) or a passive listening condition. Pure 1000 Hz sine tones were presented at three systematically varied sound intensities (40, 60 and 80 dB SPL).

The results concerning auditory evoked potentials (AEP's) and evoked GBR's obtained in the active condition predominantly corresponded to the findings of previous studies. Beside the already known augmentation of the early evoked GBR due to enhanced inter-trial phase coherence with increasing sound intensity, we also observed stronger GBR's and enhanced phase-locking in the active condition compared to passive listening, while the general shape of intensity modulation was comparable between the two conditions. In addition, phase-locking to stimulus onset was increased for stimuli of all three intensities when attended, whereas in musicians only stimuli of the highest intensity (80 dB) induced significantly increased phase-locking in the active condition.

Taken together, our results suggest that musical expertise influences attention effects on the intensity-modulated early auditory evoked GBR with respect to inter-trial phase-coherence.

Keywords
EEG; musical expertise; evoked gamma-band response; phase-locking; oscillations; attention; sound intensity; functional plasticity

Introduction

Professional musicians are a widely used and accepted model of long-term acoustic training and experience-dependent cortical plasticity. These learning-induced structural and functional alterations are rather widespread and include top-down (e.g. attention, memory) as well as bottom-up (stimulus properties) factors (for a review, see [1-4]). One important characteristic in music is its temporality, with loudness being among the most prominent parameters that vary through time. However, despite its importance, little is known about the influence of musical expertise on the processing of sound intensity.

On the other hand, the amplitude of human gamma activity (synchronous neuronal firing in the range from 30-80 Hz) has been shown to be susceptible to individual stimulus features such as size, spatial frequency and contrast [5] [6] [7] [8]. While induced gamma-band responses jitter in latency from trial to trial (non-phase-locked) and are thus cancelled out by classic averaging techniques, the evoked GBR can be detected by averaging the single trials as it is characterized by precise phase-locking to stimulus onset. In the auditory modality, a distinct dependency between stimulus intensity and evoked early gamma-band oscillations (20-130 ms) has been demonstrated [9]. In particular, these results show that an increase in sound intensity leads to enhanced inter-trial phase-locking and a corresponding augmentation of the evoked early GBR. Since the authors used an active discrimination task and gamma-band activity is known to be influenced by processing demands (e.g. [10] [11]) as well as attention (e.g. [12]), it still remains unclear whether this sound intensity dependence similarly occurs in a passive listening condition as well.

To shed some light on these open questions, we performed an EEG-study and compared early gamma-band activity in musicians and musical laymen in response to three different sound intensity levels (40dB, 60dB and 80dB SPL) in either an active forced-choice discrimination task or a passive listening condition. Based on the work of [9], we focused our analysis on the N1-P2 component of the auditory evoked potential (AEP) and the early gamma-band response (20-130 ms) recorded at the central electrode (Cz). We aimed to replicate their respective results for active tasks and additionally hypothesized to find enhanced evoked gamma-activity with increasing stimulus intensity in the passive listening condition, too. As attention has been shown to enhance early components of auditory evoked potentials (AEP), we additionally expected peak-to-peak amplitudes of the classical N1-P2 component to be stronger in general in the active condition (e.g. [13] [14]). With respect to intensity-processing related differences between the two experimental groups, we hypothesized the musician group to show a higher intensity-dependence of evoked gamma activity, leading to a stronger augmentation of the GBR with increasing stimulus intensity.

Methods

Participants

For this study, forty-three healthy volunteers with normal audiological status and no history of neurological pathology were recruited. 21 musicians (9 women, 12 men; mean age ± SD of 23.48 ± 5.49 years; 16 right-handed, 4 left-handed, 1 ambidextrous) comprised one group, with formal training starting at a mean age ± SD of 6.3 ± 1.32 years. All of them were students, music teachers and/or members of an orchestra or band and practice their instrument on a daily basis of 1-9 hours. The second group consisted of 22 musical laymen (17 women, 5 men; mean age ± SD of 25.12 ± 3.77 years;

21 right-handed, 1 ambidextrous), with no formal musical training exceeding the educational context of public elementary and secondary school. All of them gave informed written consent in accordance with procedures approved by the local ethics committee and were paid for participation.

Due to noisy data or difficulties with correctly solving the task (e.g. mixing up response buttons), the datasets of 4 musicians and 5 non-musicians had to be excluded from further analysis, leading to a total of 17 datasets per experimental group.

Stimuli and task

We applied pure 1000 Hz sinusoidal tones of three different intensities (40, 60 and 80 dB SPL) and 200 ms duration binaurally via hi-fi headphones. In the active condition, participants had to attend to the stimuli and to decide whether the just heard sine tone was of higher or lower intensity than the preceding one pressing the corresponding response-buttons on a computer keyboard with either their left or right index fingers. Allocation of the left vs. right response-buttons (lower vs. higher intensity) was counterbalanced across all subjects. In the passive condition, participants passively listened to the sine-tones and were instructed to randomly press either the right or left response-button after each sine-tone application without judging their intensity. All stimuli were presented in a fully randomized order in three blocks per condition. Each block consisted of at least 40 stimulations per sine-tone intensity, leading to a total of not less than 120 stimulations per intensity level.

AEP Recordings

Electroencephalogram was recorded using a high-density Geodesic EEG system® (GSN300; Electrical Geodesic Inc., Oregon) with 128-Channel HydroCel Geodesic Sensor Nets® (HCGSN120) (sampling rate: 500Hz; vertex electrode (Cz) as on-line recording reference; impedance < 30 kOhm). During EEG recording, participants sat in a shielded, dimly lit room and a fixation cross was presented on an LCD-screen in order to reduce eye movements.

Data Analysis - EEG, N1-P2 component analysis

EEG recordings were analysed with the BrainVision Analyzer2 software (Version 2.0.1; Brain Products GmbH, Gilching). EEG preprocessing included the following steps: band-pass filtering at 0.1-100Hz; ICA (independent component analysis) to correct for ocular artifacts (e.g. [15]); visual inspection to identify and remove trials with amplifier saturation, eye blinks, and eye movements. Sweating and muscle artifacts were excluded by means of a fully automatic raw data inspector. Further processing included: interpolation of noisy channels; re-referencing to linked mastoids; baseline correction using the signal‘s pre-stimulus portion. Finally, AEP’s for each stimulus and task were averaged for each subject and grand-averaged across subjects separately for each group.

For statistical analysis the peak-to-peak amplitude of the N1-P2 component of the auditory evoked potential (AEP) at Cz was used. The specific latency window for this component was individually identified for each subject, condition and stimulus-intensity by visual inspection. These amplitudes were then averaged separately within groups (i.e. musicians vs. non-musicians), depending on the respective condition (active vs. passive) and intensity (40 dB, 60 dB, 80 dB SPL). In the active condition, only correct trials were used for further analysis. Subsequently, a 2 x 2 x 3 repeated measure ANOVA with a between-subject factor (*group*) and two within-subjects factors (*condition* and

intensity) was computed for the N1-P2 amplitude using the multivariate approach [16]. For post-hoc tests, Bonferroni-Holm adjusted t-tests were applied for specific comparisons of significant ANOVA effects. The global level of significance was set at $p < 0.05$ for all statistical analyses.

Data Analysis - EEG, Early gamma-band responses

For the analysis of the evoked early gamma-band activity phase-locked to the stimulus, we applied a wavelet transform with a width of 3.8 cycles based on Morlet wavelets to the averaged ERP's. On the other hand, the same wavelet transformation was applied to each single trial and then the resulting wavelet transforms were averaged to analyze the non-phase-locked portion of the GBR. This measure yields the total gamma-activity consisting of both the phase-locked and non-phase-locked portion of the GBR. In addition, we also calculated the amount of inter-trial coherence, or phase-locking across trials. This procedure returns a value between 0 and 1 that determines the degree of phase-locking, with 0 indicating a high phase variability and 1 representing complete phase alignment.

As it has previously been shown, the frequency of oscillatory brain activity varies markedly between subjects (e.g. [5] [17]). Thus, we individually adapted the frequency used for the wavelet analysis by applying a separate Morlet wavelet with 12 cycles to the averaged AEP's evoked by the 80 dB tone. The individual gamma frequency was then defined as the frequency showing the highest peak in response to the 80 dB tone in a time interval between 20 and 130 ms after stimulus onset in the gamma range of that particular time-frequency representation at electrode Cz.

The peak amplitude of evoked gamma activitiy, the total gamma activity and phase-locking between 20 and 130 ms at electrode Cz defined early GBR's for the statistical analysis. Again, we computed a 2 x 2 x 3 repeated measure Analysis of Variance (ANOVA) with a between-subject factor (*group*), two within-subjects factors (*condition* and *intensity*) and used the multivariate approach to handle the problem of heteroscedasticity [15]. For specific comparisons of significant ANOVA effects, Bonferroni-Holm adjusted post hoc t-tests were calculated subsequently.

Results

Behavioral Data

All participants easily solved the task with high accuracy (mean 95.5%). Hit rates were influenced by sound intensity [$F(1, 32) = 20.995$, $p < 0.001$], with stimuli of middle intensity exhibiting a drop in accuracy (40 dB: 97.5%; 60 dB: 91.1%; 80 dB: 98.8%). As expected, hit rates in the passive condition were at chance level (mean 54.8%). The two experimental groups did not differ with respect to task performance.

EEG Data – Auditory evoked potentials (N1-P2 peak-to-peak amplitudes)

Grand-averaged waveforms evoked by the sine tones recorded at Cz and the respective peak-to-peak N1-P2 amplitudes are shown in **Figure 1A**. All stimuli elicited a prominent N1-P2 complex. Results of the 2 x 2 x 3 repeated measures ANOVA yielded a significant main effect for the factors *condition* [$F(1, 32) = 12.22$, $p = 0.001$] and *intensity* [$F(1, 32) = 26.637$, $p < 0.001$], with larger amplitudes in the active condition and for the stimuli with higher intensities. In addition, a significant *condition x intensity* interaction was found [$F(1, 32) = 4.499$, $p < 0.05$].

Subsequent post hoc comparisons indicated significant effects for 40 vs. 80 dB [$t(33) = -5.276$, $p < 0.001$] and 60 vs. 80 dB [$t(33) = -5.781$, $p < 0.001$] for the active condition. In contrast, significant effects were not only found for 40 vs. 80 dB [$t(33) = -6.822$, $p < 0.001$] and 60 vs. 80 dB [$t(33) = -6.566$, $p < 0.001$] stimuli, but also for the comparison between 40 and 60 dB stimuli [$t(33) = -2.512$, $p < 0.05$] in the passive condition. However, no group related effects were found with respect to N1-P2 peak-to-peak amplitudes.

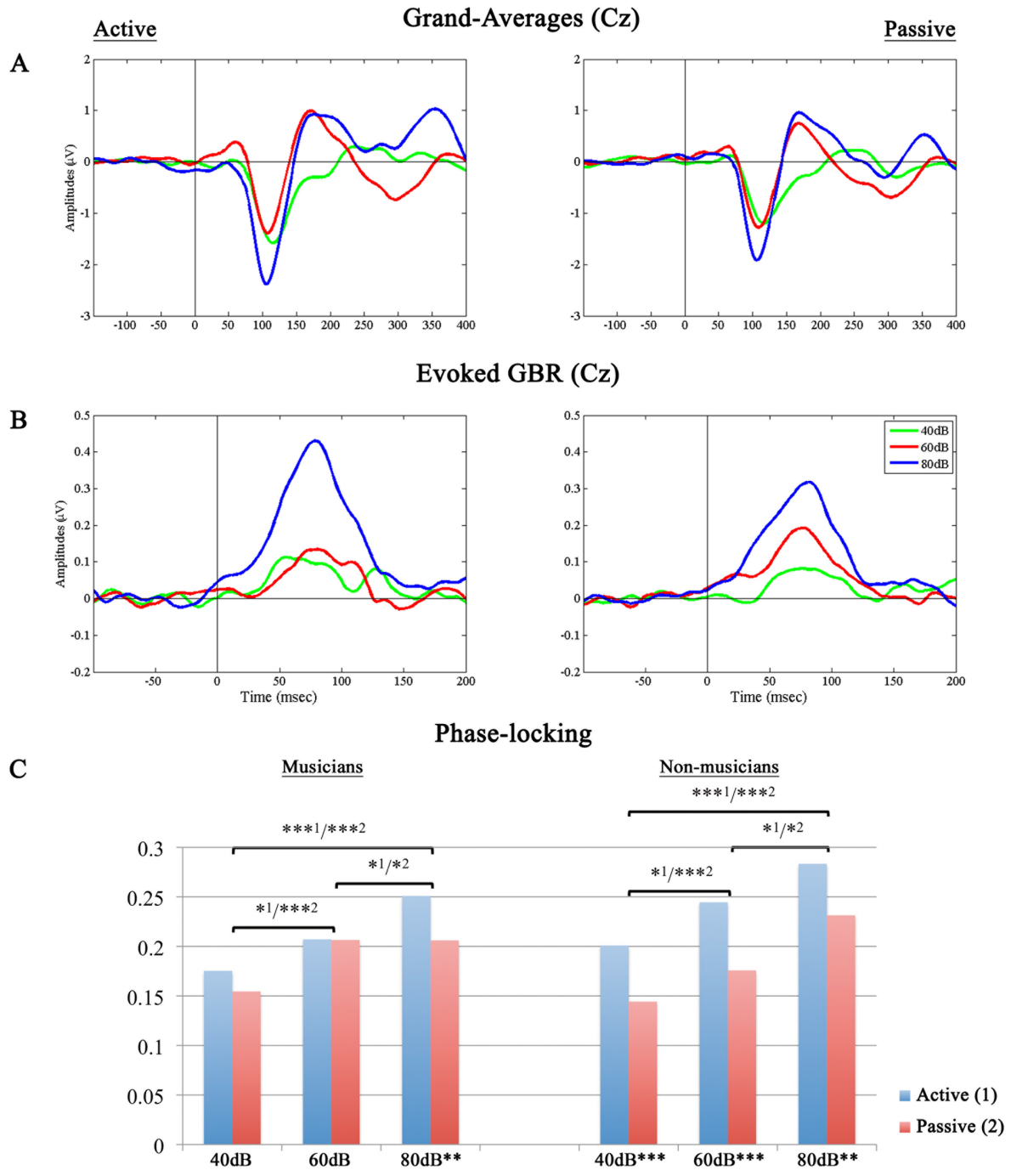

Figure 1: ***A,*** Grand-averaged waveforms over all subjects evoked by 40 (green), 60 (red) and 80 (blue) dB SPL 1000 Hz sine tones in active (left) and passive (right) conditions. ***B,*** Corresponding time courses for the evoked GBR irrespective of experimental group. ***C,*** Peak values of inter-trial phase locking for musicians (left) and non-musicians (right) in the active (1, blue) and passive (2, red) conditions. Differences indicated with brackets refer to differences independent of experimental groups (e.g. between different intensities), whereas differences indicated in the figure bar legend denote significant within-group comparisons (i.e. between conditions within the two groups). *Difference is significant at $p < 0.05$; **Difference is significant at $p < 0.01$; ***Difference is significant at $p < 0.001$.

EEG Data – Early gamma-band responses

As illustrated in **Figure 1B**, the evoked gamma-band activity increased with higher sound intensities [$F(1, 32) = 8.328$, $p = 0.001$] and amplitudes were larger in the active condition [$F(1, 32) = 8.851$, $p < 0.01$]. Post hoc t-tests revealed significant effects for 40 vs. 60 dB [$t_1(33) = -2.026$, $p < 0.05$; $t_2(33) = -2.885$, $p < 0.01$], 40 vs. 80 dB [$t_1(33) = -3.736$, $p < 0.001$; $t_2(33) = -3.483$, $p < 0.001$] and 60 vs. 80 dB [$t_1(33) = -3.118$, $p < 0.01$; $t_2(33) = -1.737$, $p < 0.05$] stimuli in both the active (t_1) and passive (t_2) condition, respectively. Moreover, evoked gamma activity was higher in the active condition for stimuli of 40 dB [$t(33) = 2.474$, $p < 0.01$] and 80 dB [$t(33) = 2.453$, $p = 0.01$] intensity.

The peak values of inter-trial phase-locking are displayed in comparison in **Figure 1C**. Here, the amount of phase-locking increased with higher intensities for the evoked GBR [$F(1, 32) = 11.741$, $p < 0.001$] and was higher in the active condition [$F(1, 32) = 47.082$, $p < 0.001$]. In addition, the ANOVA revealed a significant *condition x group* interaction [$F(1, 32) = 9.763$, $p < 0.01$]. Specific post hoc comparisons showed again significant differences between 40 and 60 dB [$t_1(33) = -2.411$, $p < 0.05$; $t_2(33) = -4.259$, $p < 0.001$], 40 and 80 dB [$t_1(33) = -3.764$, $p < 0.001$; $t_2(33) = -4.633$, $p < 0.001$] as well as 60 and 80 dB [$t_1(33) = -2.387$, $p < 0.05$, $t_2(33) = -1.712$, $p < 0.05$] stimuli in both the active (t_1) and passive (t_2) condition. In addition, inter-trial phase-coherence turned out to be stronger in the active condition in general [$t_{40dB}(16) = 4.524$, $p < 0.001$; $t_{60dB}(16) = 4.364$, $p < 0.001$; $t_{80dB}(16) = 2.735$, $p < 0.01$] only in the non-musician group. In the musician group, the difference between conditions with respect to the amount of phase-locking solely reached significance with stimuli of the highest intensity [$t_{80dB}(16) = 2.667$, $p < 0.01$].

On the contrary, the analysis of the total GBR indexing signal power, failed to reveal any significant *group*, *intensity* or *condition* effects (data not shown).

Discussion

The purpose of this study was *(a)* to examine if and to what extent intense musical training influences the processing of sound intensity and *(b)* to verify the assumption made by Schadow et al. (2007) that the intensity dependence of the N1-P2 complex and the early gamma-band response in the human EEG is also present in a fully passive listening condition.

We were able to replicate the results of [9] with respect to the intensity dependence of the N1-P2 complex and early gamma-band response in an active condition (discrimination task). Moreover, we've shown that this intensity dependence is also present in a comparable manner in a fully passive listening condition. In particular, N1-P2 peak-to-peak amplitudes significantly differed between stimuli of all three intensities in the passive condition only, whereas in the active condition, the two stimuli of lower intensity (40 and 60 dB) did not differ. Concerning early GBR, we found that increasing sound intensity also leads to an enhancement of evoked gamma-activity in the passive listening condition and that this enhancement is particularly due to an increase in inter-trial phase coherence since no increase in total GBR was observed. In addition, phase-locking to the stimulus onset in the musician group was stronger in the active condition only for stimuli with the highest intensities, whereas non-musicians showed stronger phase-locking in the active condition in general.

It is known from earlier studies (e.g. [18]) that neuronal coding of loudness is different for low and high sound intensities. In particular, intensities are signaled as changes in firing rates of neurons with similar centre frequencies up to a stimulus

intensity of about 60 dB above threshold. Intensities beyond this value are alternatively coded in such a way that the neural excitation pattern of the stimulus spreads with increasing intensity. From animal studies it is known that attention allocation to an auditory stimulus can lead to an increase in the firing rate of cortical neurons in A1 (e.g. [19]). Moreover, it has been shown that attending to a target tone induces localized increases in excitation or a reduction in inhibitory sideband near the frequency of the target in primary auditory neural assemblies [20]. Thus, the particular N1-P2 activation pattern between the active and passive condition might be caused by an attention-induced convergence of the neural response to the two lower intensity stimuli, leading to comparable N1-P2 amplitudes between 40 and 60 dB stimuli in the active condition. This would also explain the observed significant drop in accuracy for sound intensities near the firing rate/spread of excitation border (60 dB), since loudness discrimination performance could not be further improved by attention due to neural saturation at these intensities. As additional spread of neural excitation kicks in in response to stimuli beyond 60 dB above threshold, more neurons respond to the 80 dB stimulus leading to statistically distinguishable N1-P2 amplitudes compared to the two lower intensity stimuli. In contrast, neural activity is not affected by attention in the passive condition, resulting in distinct N1-P2 amplitudes in response to stimuli of all three intensities.

However, attention-induced modulation of the early gamma-band response to stimuli of different intensities observed in the present study appears somewhat different from the one just described above for AEP's. As already stated in a previous study [9], enhancement of GBR with increasing loudness of auditory stimuli is most likely caused by an increase in phase-locking with higher sound intensities. In particular, the authors suggest that this enhancement of inter-trial synchronization is due to an earlier latency of the first spike that can be recorded in response to a high intensity stimulus, while at the same time the inter-trial variability for these so-called first-spike latencies decreases, i.e. phase-locking to stimulus onset increases [21] [22]. In addition, it has been shown in a recent animal study that phase-locking to temporal stimulus properties is enhanced in auditory neurons when performing a task that requires attention [19]. Our results fit in well with these findings as intensity dependence of the GBR and phase-locking was observable in a comparable manner in both conditions, with stronger responses/increased phase-locking in the active compared to the passive task. Thus, it seems that attention also further narrows latency distributions of the first spike and/or subsequent stages of auditory processing.

We also observed attention modulation of inter-trial phase-locking to be influenced by musical proficiency. In particular, phase-locking to stimulus onset was higher in general in the active compared to the passive condition for non-musicians, whereas only phase-locking in response to stimuli of the highest intensity (80 dB) significantly differed between conditions in the musician group. Given the above described assumption about attention further narrowing spike latency distributions, it seems that intensive musical training leads to a less prominent influence of attention on the inter-trial variability of spike latencies in response to sounds of lower intensities. Maintaining a high variability of spiking latency would retain the concerned neural assemblies' flexibility to temporally code stimulus properties of sounds of low intensities, which is of particular importance in music.

Conclusion

Our results show that the processing demands of a task and/or allocation of attention differentially modulate AEP's and GBR's in response to sounds of different intensities.

This finding further supports the assumption of the two reflecting distinct and independent neuronal and possibly functional mechanisms (e.g. [9] [23]). Moreover, we've found the intensity dependence of AEP's and GBR's to be also present in a fully passive listening condition. Finally, we also observed that the influence of attention/task demands on inter-trial phase-coherence is modulated by musical proficiency. Taken together, our findings indicate that the general shape of intensity dependence of AEP's and GBR's is similar in active and passive conditions, but the strength of both AEP and GBR amplitudes tends to be augmented in an active discrimination task. Furthermore, musical expertise shapes attention effects on intensity-modulated inter-trial phase-coherence in early auditory evoked gamma-band responses.

Acknowledgement:

This study was supported by a grant of the Swiss National Science Foundation (SNF) to Lutz Jäncke (32003B_121927).

References

1. Jäncke L Music drives brain plasticity. *F1000 Biol Rep* 2009; 1-6.

2. Münte TF, Altenmüller E, Jäncke L The musician's brain as a model of neuroplasticity. *Nat Rev Neurosci* 2002; **3**: 473-478.

3. Schlaug G The brain of musicians. A model for functional and structural adaptation. *Annals of the New York Academy of Sciences* 2001; **930**: 281-299.

4. Shahin AJ Neurophysiological influence of musical training on speech perception. *Front Psychol* 2011; **2**: 126.

5. Busch NA, Debener S, Kranczioch C, Engel AK, Herrmann CS Size matters: effects of stimulus size, duration and eccentricity on the visual gamma-band response. *Clin Neurophysiol* 2004; **115**: 1810-1820.

6. Tzelepi A, Bezerianos T, Bodis-Wollner I Functional properties of sub-bands of oscillatory brain waves to pattern visual stimulation in man. *Clin Neurophysiol* 2000; **111**: 259-269.

7. Fründ I, Busch NA, Schadow J, Körner U, Herrmann CS From perception to action: phase-locked gamma oscillations correlate with reaction times in a speeded response task. *BMC Neurosci* 2007; **8**: 27.

8. Schadow J, Lenz D, Thaerig S, Busch NA, Fründ I, Rieger JW *et al.* Stimulus intensity affects early sensory processing: visual contrast modulates evoked gamma-band activity in human EEG. *Int J Psychophysiol* 2007; **66**: 28-36.

9. Schadow J, Lenz D, Thaerig S, Busch NA, Fründ I, Herrmann CS Stimulus intensity affects early sensory processing: sound intensity modulates auditory evoked gamma-band activity in human EEG. *Int J Psychophysiol* 2007; **65**: 152-161.

10. Senkowski D, Herrmann CS Effects of task difficulty on evoked gamma activity and ERPs in a visual discrimination task. *Clin Neurophysiol* 2002; **113**: 1742-1753.

11. Simos PG, Papanikolaou E, Sakkalis E, Micheloyannis S Modulation of gamma-band spectral power by cognitive task complexity. *Brain Topogr* 2002; **14**: 191-196.

12. Busch NA, Schadow J, Fründ I, Herrmann CS Time-frequency analysis of target detection reveals an early interface between bottom-up and top-down processes in the gamma-band. *NeuroImage* 2006; **29**: 1106-1116.

13. Lange K, Rösler F, Röder B Early processing stages are modulated when auditory stimuli are presented at an attended moment in time: an event-related potential study. *Psychophysiology* 2003; **40**: 806-817.

14. Woldorff MG, Gallen CC, Hampson SA, Hillyard SA, Pantev C, Sobel D *et al.* Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proc Natl Acad Sci USA* 1993; **90**: 8722-8726.

15. Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin Neurophysiol* 2000; **111**: 1745-1758.

16. O'Brien RG, Kaiser MK MANOVA method for analyzing repeated measures designs: an extensive primer. *Psychol Bull* 1985; **97**: 316-333.

17. Zaehle T, Fründ I, Schadow J, Thärig S, Schoenfeld MA, Herrmann CS Inter- and intra-individual covariations of hemodynamic and oscillatory gamma responses in the human cortex. *Frontiers in human neuroscience* 2009; **3**: 8.

18. Jäncke L, Shah NJ, Posse S, Grosse-Ryuken M, Müller-Gärtner HW Intensity coding of auditory stimuli: an fMRI study. *Neuropsychologia* 1998; **36**: 875-883.

19. Niwa M, Johnson JS, O'Connor KN, Sutter ML Active engagement improves primary auditory cortical neurons' ability to discriminate temporal modulation. *J Neurosci* 2012; **32**: 9323-9334.

20. Fritz J, Shamma S, Elhilali M, Klein D Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci* 2003; **6**: 1216-1223.

21. Heil P, Irvine DR First-spike timing of auditory-nerve fibers and comparison with auditory cortex. *Journal of Neurophysiology* 1997; **78**: 2438-2454.

22. Heil P First-spike latency of auditory neurons revisited. *Curr Opin Neurobiol* 2004; **14**: 461-467.

23. Bertrand O, Tallon-Baudry C Oscillatory gamma activity in humans: a possible role for object representation. *Int J Psychophysiol* 2000; **38**: 211-223.

4. General Discussion

There is growing evidence that musical expertise has a substantial impact on the anatomical and functional organization of the human brain. As a consequence, the brain of musicians serves as a favored model for neuroplasticity. In this context, the aim of the present doctoral thesis was to investigate if and to what extent intense musical training also leads to functional alterations in early stages of speech-related auditory processing. To this end I designed and performed three different studies focussing on three specific research topics particularly involving speech-related low-level auditory processing. To accomplish this goal, I decided to use high-density electroencephalography in combination with classical component analysis as well as more sophisticated methods to evaluate the EEG data obtained in all three studies. In the following section i will first compendiously summarize the main results of the published / submitted original research articles and subsequently discuss them in a broader context by highlighting common patterns as well as possible underlying mechanisms of the observed cross-domain functional neuroplasticity.

Experiment 1 was designed to determine whether neurophysiological processing of the voice-onset time (VOT) parameter in acoustic stimuli and particularly in speech is altered in musicians compared to musical laymen. For this purpose, we presented consonant-vowel (CV) syllables and corresponding noise-analogs with varying voice-onset times ranging from approximately 04 ms up to 59 ms. The neurophysiological responses of 26 German speaking musicians and non-musicans to these stimuli were analyzed and compared using EEG in combination with classical AEP component analysis as well as the EP-mapping approach. The results showed that musicians process unvoiced acoustic stimuli (i.e. stimuli with long VOT's) differently than musical laymen, irrespective of whether these stimuli are speech or non-speech. In conclusion, musicians seem to *a)* utilize the same network to analyze unvoiced stimuli as they use for the analysis of voiced stimuli and *b)* to devote more neurophysiological resources into the analysis of unvoiced segments.

In **experiment 2** we were specifically interested in whether possible musical training induced alterations in the neural underpinnings of motor-induced suppression (MIS) of cortical responses to self-initiated sounds provided by internally operating forward models are restricted to the sounds to which musicians are mainly exposed during their

daily practice routines or if they are more widespread and are also affecting the processing of speech. Moreover, we tried to replicate and verify our results from experiment 1 by using a subset of the speech stimuli (CV-syllables) we already used in said experiment. In particular, we compared cortical responses to predictable and unpredictable self- vs. externally initiated zero-delay CV-syllables and piano tones between musicians and musical laymen. Again, we used the AEP as well as the topographic pattern analysis (EP-mapping/microstate analysis) approach. Taken together, the results from this particular experiment suggest that besides the already known effect of motor-induced suppression of cortical responses to self-triggered sounds, internally operating forward models also facilitate early acoustic analysis of complex tones by means of faster processing time in the first ~200 ms after stimulus onset. More importantly, musicians seem to profit from this facilitation also during the analysis of speech sounds whereas musical laymen do not show such a cross-domain facilitation-effect. Finally, we were not able to replicate our results from experiment 1, most likely due to substantial differences regarding experimental groups and tasks.

In the end, **experiment 3** was intended to reveal possible differences between musicians and non-musicians with respect to the processing of sound intensity. Furthermore, we aimed to replicate the results of a similiar former study of another research group (Schadow et al., 2007), and complement their findings by introducing a passive listening condition in comparison to an active discrimination task. Here, we binaurally presented pure 1000 Hz sine tones at three different intensities (40, 60 and 80 dB SPL) and compared the resulting N1-P2 AEP complex as well as early auditory gamma-band response (GBR) of 17 musicians to those of 17 musical laymen obtained during either a forced-choice discrimination task (active) or a fully passive listening condition. Doing so, we were not only able to replicate the results of (Schadow et al., 2007), but also to provide evidence that the general shape of intensity dependence of the early evoked GBR and particularly the inter-trial phase variability is comparable in active and passive conditions. Though, GBR's were generally stronger and inter-trial phase-locking was increased in the active compared to the passive condition. In addition, inter-trial phase coherence was increased for stimuli of all three intensities when attended in non-musicians, whereas in musicians only stimuli of the highest intensity (80 dB SPL) induced significantly increased phase-locking in the active condition. In sum, experiment 3 showed that musical expertise influences attention

effects on the intensity-modulated early auditory evoked GBR with respect to inter-trial phase-coherence.

A common ground of all three experiments is that they provide further evidence for the notion of shared neuronal networks and computational principles between music and speech to begin with. As already noted in the introduction section (chapter 1.2.3.), musical training has to target and alter the particular neuronal assemblies that also underlie elementary speech processing in order to transfer modified auditory skills to basic speech functions. That brain regions involved in basic musical and speech processing indeed overlap can readily be seen by comparing two exemplary fMRI studies investigating cortical responses to speech (Jäncke et al., 2002) and melodic sounds (Patterson et al., 2002), respectively. Both studies identified the planum temporale (PT) to be particularly activated in response to the respective stimuli they used in their experiments (i.e. CV-syllables / vowels and melodic sounds). Similar findings have been reported by Rogalsky et al. (2011), who found overlapping bilateral activation patterns in auditory-related supratemporal cortex for both sentences and melodies. Taken together, these results clearly demonstrate that music and speech sounds at least partially drive overlapping neuronal assemblies in auditory-related cortical areas. Given this matched prerequisite for transfer effects from musical training to elementary speech processing to occur, it comes as no surprise that the results obtained in the original research series within the framework of this doctoral thesis revealed transfer from musical proficiency to different realms of speech processing. In particular, functional plasticity as a function of musical expertise occured with respect to extraction and analysis of elementary phonetic cues, sensory-motor interaction as well as higher-order cognition (i.e. attention).

These results fit in well with the fact that the human auditory system is composed of specialized and interacting multi-staged networks encoding and analyzing spatial, spectral and temporal information while at the same time, music and speech are perceptually distinct but share many commonalities at both an acoustic and cognitive level. Accordingly, neuroplastic changes due to musical training have already been repeatedly shown to occur on different stages of auditory processing just as in the work at hand, suggesting miscellaneous underlying mechanisms of such plasticity.

First, enhanced encoding of sound properties in musicians can already be observed at the level of the brainstem (e.g. Musacchia et al., 2007; Parbery-Clark et al., 2011b). Incoming sound travels from the cochlea to the auditory cortex through a series of brainstem nuclei that extract and process sound information. This pathway is accompagnied by the so-called cortico-fugal network, which basically represents a system of feedback pathways that connects the cortex to the brainstem and the cochlea in a top-down manner. Since neuroplastic changes as a function of musical expertise have been observed in the auditory cortex as well as in lower-level sensory regions, the enhanced subcortical encoding of stimulus features observed in musicians is possibly a result of strengthened top-down feedback pathways. Thus, the ability to rapidly detect, sequence and encode specific sound patterns in subcortical nuclei appears to be improved as a consequence of active engagement with music, which in turn enables the cortex to selectively enhance predictable features of the auditory signal already at the level of the brainstem (Kraus and Chandrasekaran, 2010). Even though the doctoral thesis at hand does not focus on transfer effects and plasticity in subcortical regions per se, it is nonetheless important to bear in mind that the subcortical projections to the auditory cortex of musicians can carry an already altered signal, thereby contributing to differences ostensibly appearing at later levels of processing. Hence, future research should be conducted on this particular topic in order to further disentangle the interdependency of musical training-induced plasticity in subcortical and cortical regions and / or networks.

Second, as the results of experiment 2 clearly demonstrate, musical training exerts its influence not only on motor and auditory systems separately but also on their interactions. Since the two are closely linked together during musical performance, this result is actually not surprising, and corresponding findings have already repeatedly been shown in the past (e.g. Bangert and Altenmüller, 2003; Bangert et al., 2006; D'Ausilio et al., 2006; Lahav et al., 2007). Complementary results have been reported by (Lappe et al., 2008), showing that multimodal musical training can induce even stronger neuroplastic changes in auditory processing compared to unimodal training. Taken together, these results readily highlight the existence of strong functional interconnections between the auditory and motor system during the perception and production of music. Given this strong functional interconnection, it can be deducted that Hebbian mechanisms based on simultaneous inputs leading to changes in synaptic strength appear to be a prime factor most likely responsible for these multimodal plastic

effects (Herholz and Zatorre, 2012). Corroborating evidence for this assumption comes from a study using transcranial magnetic stimulation (TMS), showing augmented motor excitability in pianists during auditory presentation of a rehearsed but not of a non-rehearsed musical piece (D'Ausilio et al., 2006). In addition, two studies using the same method indicate that simultaneous activation of cortical areas by a stimulus input and by a TMS pulse results in local functional plasticity (Stefan et al., 2000; Stefan et al., 2002). In particular, the thresholds for motor evoked potentials (MEP) by TMS are modulated as a function of the delay between the stimulus input and the TMS pulse. Moreover, the increase of single-pulse TMS-evoked MEP's was prevented when the simultaneous activations were conducted under the influence of dextromethorphan, an N-methyl-D-aspartate (NMDA) antagonist known to block long-term potentiation (LTP). Hence, these processes can be interpreted as analoguous to LTP and LTD on a cellular level (Herholz and Zatorre, 2012; Hoogendam et al., 2010).

Third, it is known that top-down influences, particularly attention and the reward value of a task, can exert a substantial impact on neuroplastic processes in auditory-related cortex (e.g. David et al., 2012; Fritz et al., 2007; Polley et al., 2006). Insofar, musical training represents an ideal mediator of neuroplasticity, as it requires active engagement with musical sounds and their production in a highly concentrated manner, while at the same time the successful performance of pretentious musical pieces can be extremely rewarding. On the other hand, the results of our third study suggest that the relationship between musical training and top-down processes is mutual, at least with respect to attention. This assumption receives support from a recent study by (Strait and Kraus, 2011), showing strengthened brain networks for selective auditory attention in musicians indicated by decreased prefrontal response variability to speech sounds with auditory attention. In an earlier study, the same group was able to demonstrate lower perceptual thresholds in musicians specifically for auditory tasks that relate to cognitive abilities such as backward masking and auditory attention (Strait et al., 2010). Other researchers reported training-induced changes in attentional skills as reflected by enhanced short-term plasticity of the P3a/P3b AEP components in musicians (Seppänen et al., 2012). Taken together, these findings substantiate that musical training also leads to functional neuroplasticity on higher cognitive levels. This is of particular importance when considering the fact that long-term musical training seems not only to be capable of altering specific structural and functional properties of the brain, but also to amplify the potential for new short-term learning and plasticity (e.g. Herholz et al., 2011;

Rosenkranz et al., 2007). Since it is still not clear whether such enhancement is due to top-down influences such as attention to relevant input and / or to an already altered sensory system on lower levels of processing, taking said observations into account may further elucidate the origins of this long-term training related phenomenon. Taking into account that attention substantially influences neuroplastic changes due to training as stated above, alterations in attentional functions may therefore also contribute to the augmented potential for short-term plasticity observed in musicians.

In conclusion, the results reported and explicated in this doctoral thesis are well in line with pre-existing work on the topic of functional neuroplasticity in the context of musical expertise (for a review, consider e.g. Herholz and Zatorre, 2012; Jäncke, 2009b; Shahin, 2011). They further elucidate the massive impact of long-term musical training on different stages of auditory processing as well as auditory-motor interaction and higher-order cognitive functions. As an additional point I also demonstrated that functional plasticity induced by long-term musical training is not restricted to music, but readily transfers into the domain of subsegmental spectro-temporal speech processing. However, i want to emphasize that despite the myriad of studies conducted to investigate the neuronal fundamentals of the impressive skills held by professional musicians, it still remains to be fully described how all the neuronal devices involved in playing music work together and thereby enable steady structural and functional cross-domain reorganization in the brain.

5. Abbreviations

A1: Primary auditory cortex

AAHC: Atomize and Agglomerate Hierarchical Clustering

AEP: Auditory evoked potential

BOLD: Blood oxygen level dependent

CMA: Cingulate motor area

CV: Consonant-vowel

Cz: Central electrode / Vertex electrode

dB: Decibel

DNA: Deoxyribonucleic acid

DTI: Diffusion tensor imaging

EEG: Electroencephalography

EP: Evoked potential

ERO: Event-related oscillation

ERP: Event-related potential

FFR: Frequency following response

fMRI: Functional magnetic resonance imaging

f0/F0: Fundamental frequency

GBR: Gamma band response

GEV: Global explained variance

LTD: Long-term depression

LTP: Long-term potentiation

MEP: Motor evoked potential

MLR: Mid-latency response

MIS: Motor induced suppression

mRNA: Messenger ribonucleic acid

M1: Primary motor area

NMDA: N-methyl-D-aspartate

NPAC: Non-primary auditory cortex

Pre-SMA: Pre-supplementary motor area

PT: Planum temporale

7. References

Ahissar, E., Abeles, M., Ahissar, M., Haidarliu, S., & Vaadia, E. (1998). Hebbian-like functional plasticity in the auditory cortex of the behaving monkey. *Neuropharmacology*, *37*(4-5), 633-655.

Atcherson, S. R., Gould, H. J., Pousson, M. A., & Prout, T. M. (2006). Long-term stability of N1 sources using low-resolution electromagnetic tomography. *Brain Topogr*, *19*(1-2), 11-20.

Bangert, M., & Altenmüller, E. O. (2003). Mapping perception to action in piano practice: a longitudinal DC-EEG study. *BMC neuroscience*, *4*, 26.

Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H. et al. (2006). Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. *Neuroimage*, *30*(3), 917-926.

Bangert, M., & Schlaug, G. (2006). Specialization of the specialized in features of external human brain morphology. *Eur J Neurosci*, *24*(6), 1832-1834.

Bao, S., Chang, E. F., Woods, J., & Merzenich, M. M. (2004). Temporal plasticity in the primary auditory cortex induced by operant perceptual learning. *Nat Neurosci*, *7*(9), 974-981.

Bavelier, D., Green, C. S., Pouget, A., & Schrater, P. (2012). Brain plasticity through the life span: learning to learn and action video games. *Annu Rev Neurosci*, *35*, 391-416.

Bermudez, P., Lerch, J. P., Evans, A. C., & Zatorre, R. J. (2009). Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cereb Cortex*, *19*(7), 1583-1596.

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011a). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of cognitive neuroscience*, *23*(2), 425-434.

Bidelman, G. M., Krishnan, A., & Gandour, J. T. (2011b). Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *Eur J Neurosci*, *33*(3), 530-538.

Bliss, T. V., & Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J Physiol (Lond)*, *232*(2), 331-356.

Bosnyak, D. J., Eaton, R. A., & Roberts, L. E. (2004). Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. *Cereb Cortex*, *14*(10), 1088-1099.

Brown, S., Martinez, M. J., & Parsons, L. M. (2006). Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *Eur J Neurosci*, *23*(10), 2791-2803.

Buschkuehl, M., Jaeggi, S. M., & Jonides, J. (2012). Neuronal effects following working memory training. *Dev Cogn Neurosci*, *2 Suppl 1*, S167-79.

Cooke, S. F., & Bliss, T. V. (2006). Plasticity in the human central nervous system. *Brain*, *129*(Pt 7), 1659-1673.

D'Ausilio, A., Altenmüller, E., Olivetti Belardinelli, M., & Lotze, M. (2006). Cross-modal plasticity of the motor cortex while listening to a rehearsed musical piece. *Eur J Neurosci*, *24*(3), 955-958.

David, S. V., Fritz, J. B., & Shamma, S. A. (2012). Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proc Natl Acad Sci USA*, *109*(6), 2144-2149.

de Villers-Sidani, E., Simpson, K. L., Lu, Y. F., Lin, R. C., & Merzenich, M. M. (2008). Manipulating critical period closure across different sectors of the primary auditory cortex. *Nat Neurosci*, *11*(8), 957-965.

Douglas, R. M., & Goddard, G. V. (1975). Long-term potentiation of the perforant path-granule cell synapse in the rat hippocampus. *Brain Res*, *86*(2), 205-215.

Draganski, B., & May, A. (2008). Training-induced structural changes in the adult human brain. *Behav Brain Res*, *192*(1), 137-142.

Elmer, S. (2010). *Functional and Anatomical Correlates of Language Expertise.* Unpublished Doctoral Thesis, University of Zurich, Zurich.

Elmer, S., Meyer, M., & Jäncke, L. (2012). Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cereb Cortex*, *22*(3), 650-658.

Engvig, A., Fjell, A. M., Westlye, L. T., Moberget, T., Sundseth, Ø., Larsen, V. A. et al. (2010). Effects of memory training on cortical thickness in the elderly. *Neuroimage*, *52*(4), 1667-1676.

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., & Gibson, E. (2009). Structural integration in language and music: evidence for a shared system. *Mem Cognit*, *37*(1), 1-9.

François, C., Tillmann, B., & Schön, D. (2012). Cognitive and methodological considerations on the effects of musical expertise on speech segmentation. *Annals of the New York Academy of Sciences*, *1252*, 108-115.

Fritz, J., Elhilali, M., & Shamma, S. (2005). Active listening: task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hear Res*, *206*(1-2), 159-176.

Fritz, J. B., Elhilali, M., & Shamma, S. A. (2007). Adaptive changes in cortical receptive fields induced by attention to complex sounds. *Journal of Neurophysiology*, *98*(4), 2337-2346.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of cognitive neuroscience*, *16*(6), 1010-1021.

Gao, Z., van Beugen, B. J., & De Zeeuw, C. I. (2012). Distributed synergistic plasticity and cerebellar learning. *Nat Rev Neurosci*, *13*(9), 619-635.

Gaser, C., & Schlaug, G. (2003a). Gray matter differences between musicians and nonmusicians. *Ann N Y Acad Sci*, *999*, 514-517.

Gaser, C., & Schlaug, G. (2003b). Brain structures differ between musicians and non-musicians. *J Neurosci*, *23*(27), 9240-9245.

Gazzaniga, M. S., Ivry, R. B., & Mangun, G. R. (2002). Learning and Memory. In J. W. Durbin (Ed.), *Cognitive Neuroscience* (2 ed., pp. 301-350). New York, NY 10110: W. W. Norton & Company Inc. 500 Fifth Avenue.

Hackett, T. A., Preuss, T. M., & Kaas, J. H. (2001). Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J Comp Neurol*, *441*(3), 197-222.

Hall, J. W. (2006). *New Handbook of Auditory Evoked Responses.* Boston: Pearson.

Hannon, E. E., Snyder, J. S., Eerola, T., & Krumhansl, C. L. (2004). The role of melodic and temporal cues in perceiving musical meter. *J Exp Psychol Hum Percept Perform*, *30*(5), 956-974.

Hebb, D. O. (1949). *The organization of behavior.* New York: Wiley.

Henze, D. A., Borhegyi, Z., Csicsvari, J., Mamiya, A., Harris, K. D., & Buzsáki, G. (2000). Intracellular features predicted by extracellular recordings in the hippocampus in vivo. *Journal of Neurophysiology*, *84*(1), 390-400.

Herholz, S. C., Boh, B., & Pantev, C. (2011). Musical training modulates encoding of higher-order regularities in the auditory cortex. *Eur J Neurosci*, *34*(3), 524-529.

Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: behavior, function, and structure. *Neuron*, *76*(3), 486-502.

Herrmann, C., Grigutsch, M., & Busch, N. A. (2005). EEG oscillations and wavelet analysis. *Event-related potentials: A methods handbook,* 229-259.

Hertrich, I., Mathiak, K., Lutzenberger, W., & Ackermann, H. (2000). Differential impact of periodic and aperiodic speech-like acoustic signals on magnetic M50/M100 fields. *Neuroreport*, *11*(18), 4017-4020.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, *182*(4108), 177-180.

Hoogendam, J. M., Ramakers, G. M., & Di Lazzaro, V. (2010). Physiology of repetitive transcranial magnetic stimulation of the human brain. *Brain Stimul*, *3*(2), 95-118.

Hutchinson, S., Lee, L. H., Gaab, N., & Schlaug, G. (2003). Cerebellar volume of musicians. *Cereb Cortex*, *13*(9), 943-949.

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C. et al. (2009). Musical training shapes structural brain development. *Journal of Neuroscience*, *29*(10), 3019-3025.

Jäncke, L. (2009a). The plastic human brain. *Restor Neurol Neurosci*, *27*(5), 521-538.

Jäncke, L. (2009b). Music drives brain plasticity. *F1000 Biol Rep,* 1-6.

Jäncke, L., Shah, N. J., & Peters, M. (2000). Cortical activations in primary and secondary motor areas for complex bimanual movements in professional pianists. *Brain Res Cogn Brain Res*, *10*(1-2), 177-183.

Jäncke, L., Wüstenberg, T., Scheich, H., & Heinze, H. J. (2002). Phonetic perception and the temporal cortex. *Neuroimage*, *15*(4), 733-746.

Kleber, B., Birbaumer, N., Veit, R., Trevorrow, T., & Lotze, M. (2007). Overt and imagined singing of an Italian aria. *Neuroimage*, *36*(3), 889-900.

Koelsch, S., Gunter, T., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing: "nonmusicians" are musical. *Journal of cognitive neuroscience*, *12*(3), 520-541.

Koeneke, S., Lutz, K., Wüstenberg, T., & Jäncke, L. (2004). Long-term training affects cerebellar processing in skilled keyboard players. *Neuroreport*, *15*(8), 1279-1282.

Koenig, T., & Pascual-Marqui, R. D. (2009). Multichannel frequency and time-frequency analysis. In C. M. Michel, T. Koenig, D. Brandeis, & L. R. R. Gianotti (Eds.), *Electrical Neuroimaging* (pp. 145-168). Cambridge: Cambridge University Press.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat Rev Neurosci*, *11*(8), 599-605.

Krishnan, A., Swaminathan, J., & Gandour, J. T. (2009). Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *Journal of cognitive neuroscience*, *21*(6), 1092-1105.

Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *J Neurosci*, *27*(2), 308-314.

Lappe, C., Herholz, S. C., Trainor, L. J., & Pantev, C. (2008). Cortical plasticity induced by short-term unimodal and multimodal musical training. *J Neurosci*, *28*(39), 9632-9639.

Lappe, C., Trainor, L. J., Herholz, S. C., & Pantev, C. (2011). Cortical plasticity induced by short-term multimodal musical rhythm training. *PloS one*, *6*(6), e21493.

Lehmann, D. (1987). Principles of spatial analysis. In A. S. Gevins & A. Redmond (Eds.), *Handbook of electroencephalography and clinical neurophysiology, vol. 1: methods of analysis of brain electrical and magnetic signals* (pp. 309-354). Amsterdam: Elsevier.

Lømo, T. (1966). Frequency potentiation of excitatory synaptic activity in the dentate area of the hippocampal formation. *Acta Physiologica Scandinavica*, *68*(Suppl 227), 128.

Lømo, T. (2003). The discovery of long-term potentiation. *Philos Trans R Soc Lond, B, Biol Sci*, *358*(1432), 617-620.

Lotze, M., Scheler, G., Tan, H. R. M., Braun, C., & Birbaumer, N. (2003). The musician's brain: functional imaging of amateurs and professionals during performance and imagery. *Neuroimage*, *20*(3), 1817-1829.

Lövdén, M., Bodammer, N. C., Kühn, S., Kaufmann, J., Schütze, H., Tempelmann, C. et al. (2010). Experience-dependent plasticity of white-matter microstructure extends into old age. *Neuropsychologia*, *48*(13), 3878-3883.

Magne, C., Schön, D., & Besson, M. (2003). Prosodic and melodic processing in adults and children. Behavioral and electrophysiologic approaches. *Ann N Y Acad Sci*, *999*, 461-476.

Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *Journal of cognitive neuroscience*, *18*(2), 199-211.

Marie, C., Magne, C., & Besson, M. (2011). Musicians and the metric structure of words. *Journal of cognitive neuroscience*, *23*(2), 294-305.

Meyer, M., Elmer, S., Ringli, M., Oechslin, M. S., Baumann, S., & Jancke, L. (2011). Long-term exposure to music enhances the sensitivity of the auditory system in children. *Eur J Neurosci*, *34*(5), 755-765.

Michel, C. M., Koenig, T., & Brandeis, D. (2009). "Electrical neuroimaging in the time domain - microstate analysis". In C. M. Michel, T. Koenig, D. Brandeis, L. R. R. Gianotti, & J. Wackermann (Eds.), *Electrical Neuroimaging* (pp. 123-125). Cambridge: Cambridge University Press.

Murray, M. M., Brunet, D., & Michel, C. M. (2008). Topographic ERP analyses: a step-by-step tutorial review. *Brain Topogr*, *20*(4), 249-264.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc Natl Acad Sci USA*, *104*(40), 15894-15898.

Musacchia, G., Strait, D., & Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hear Res*, *241*(1-2), 34-42.

Neuhaus, C., & Knösche, T. R. (2008). Processing of pitch and time sequences in music. *Neurosci Lett*, *441*(1), 11-15.

Ohl, F. W., & Scheich, H. (2005). Learning-induced plasticity in animal and human auditory cortex. *Curr Opin Neurobiol*, *15*(4), 470-477.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*, *392*(6678), 811-814.

Parbery-Clark, A., Strait, D., Anderson, S., Hittner, E., & Kraus, N. (2011a). Musical experience and the aging auditory system: implications for cognitive abilities and hearing speech in noise. *PLoS ONE*, *6*(5), e18082.

Parbery-Clark, A., Strait, D. L., & Kraus, N. (2011b). Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. *Neuropsychologia*, *49*(12), 3338-3345.

Patel, A. D. (2003). Rhythm in language and music: parallels and differences. *Annals of the New York Academy of Sciences*, *999*, 140-143.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, *36*(4), 767-776.

Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annu Rev Psychol*, *56*, 89-114.

Polley, D. B., Steinberg, E. E., & Merzenich, M. M. (2006). Perceptual learning directs auditory cortical map reorganization through top-down influences. *J Neurosci*, *26*(18), 4970-4982.

Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., & La Mantia, A. S. (2008). *Neuroscience* (4 ed.). Sinauer Associates Inc., U.S.

Ramachandran, V. S., & Rogers-Ramachandran, D. (2000). Phantom limbs and neural plasticity. *Arch Neurol*, *57*(3), 317-320.

Ramachandran, V. S., Rogers-Ramachandran, D., & Cobb, S. (1995). Touching the phantom limb. *Nature*, *377*(6549), 489-490.

Ramachandran, V. S., Rogers-Ramachandran, D., & Stewart, M. (1992). Perceptual correlates of massive cortical reorganization. *Science*, *258*(5085), 1159-1160.

Recanzone, G. H., Schreiner, C. E., & Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *J Neurosci*, *13*(1), 87-103.

Reinke, K. S., He, Y., Wang, C., & Alain, C. (2003). Perceptual learning modulates sensory evoked response during vowel segregation. *Brain Res Cogn Brain Res*, *17*(3), 781-791.

Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *J Neurosci*, *31*(10), 3843-3852.

Rosenkranz, K., Williamon, A., & Rothwell, J. C. (2007). Motorcortical excitability and synaptic plasticity is enhanced in professional musicians. *J Neurosci*, *27*(19), 5200-5206.

Sammler, D., Koelsch, S., Ball, T., Brandt, A., Elger, C. E., Friederici, A. D. et al. (2009). Overlap of musical and linguistic syntax processing: intracranial ERP evidence. *Annals of the New York Academy of Sciences*, *1169*, 494-498.

Schadow, J., Lenz, D., Thaerig, S., Busch, N. A., Fründ, I., & Herrmann, C. (2007). Stimulus intensity affects early sensory processing: sound intensity modulates auditory evoked gamma-band activity in human EEG. *Int J Psychophysiol*, *65*(2), 152-161.

Schlaug, G., Forgeard, M., Zhu, L., Norton, A., Norton, A., & Winner, E. (2009). Training-induced neuroplasticity in young children. *Annals of the New York Academy of Sciences*, *1169*, 205-208.

Schlaug, G., Jäncke, L., Huang, Y., Staiger, J. F., & Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia*, *33*(8), 1047-1055.

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat Neurosci*, *5*(7), 688-694.

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J. et al. (2005). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nat Neurosci*, *8*(9), 1241-1247.

Seppänen, M., Pesonen, A. K., & Tervaniemi, M. (2012). Music training enhances the rapid plasticity of P3a/P3b event-related brain potentials for unattended and attended target sounds. *Atten Percept Psychophys*, *74*(3), 600-612.

Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J Neurosci*, *23*(13), 5545-5552.

Shahin, A., Roberts, L. E., & Trainor, L. J. (2004). Enhancement of auditory cortical development by musical experience in children. *Neuroreport*, *15*(12), 1917-1921.

Shahin, A. J. (2011). Neurophysiological influence of musical training on speech perception. *Front Psychol*, *2*, 126.

Shahin, A. J., Roberts, L. E., Chau, W., Trainor, L. J., & Miller, L. M. (2008). Music training leads to the development of timbre-specific gamma band activity. *NeuroImage*, *41*(1), 113-122.

Shetake, J. A., Engineer, N. D., Vrana, W. A., Wolf, J. T., & Kilgard, M. P. (2012). Pairing tone trains with vagus nerve stimulation induces temporal plasticity in auditory cortex. *Exp Neurol*, *233*(1), 342-349.

Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A., & Roberts, N. (2002). Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. *Neuroimage*, *17*(3), 1613-1622.

Song, J. H., Skoe, E., Wong, P. C., & Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of cognitive neuroscience*, *20*(10), 1892-1902.

Stefan, K., Kunesch, E., Benecke, R., Cohen, L. G., & Classen, J. (2002). Mechanisms of enhancement of human motor cortex excitability induced by interventional paired associative stimulation. *J Physiol (Lond)*, *543*(Pt 2), 699-708.

Stefan, K., Kunesch, E., Cohen, L. G., Benecke, R., & Classen, J. (2000). Induction of plasticity in the human motor cortex by paired associative stimulation. *Brain*, *123 Pt 3*, 572-584.

Strait, D., & Kraus, N. (2011). Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Front Psychol*, *2*, 113.

Strait, D., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hear Res*, *261*(1-2), 22-29.

Strait, D., Parbery-Clark, A., Hittner, E., & Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain Lang*, *123*(3), 191-201.

Tashiro, A., & Yuste, R. (2003). Structure and molecular organization of dendritic spines. *Histol Histopathol*, *18*(2), 617-634.

Toiviainen, P., & Krumhansl, C. L. (2003). Measuring and modeling real-time responses to music: the dynamics of tonality induction. *Perception*, *32*(6), 741-766.

Trainor, L. J., Shahin, A., & Roberts, L. E. (2003). Effects of musical training on the auditory cortex in children. *Annals of the New York Academy of Sciences*, *999*, 506-513.

Tremblay, K., & Kraus, N. (2002). Auditory training induces asymmetrical changes in cortical neural activity. *J Speech Lang Hear Res*, *45*(3), 564-572.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear and hearing*, *22*(2), 79-90.

Wan, C. Y., & Schlaug, G. (2010). Music making as a tool for promoting brain plasticity across the life span. *Neuroscientist*, *16*(5), 566-577.

Wilson, S. J., Abbott, D. F., Lusher, D., Gentle, E. C., & Jackson, G. D. (2011). Finding your voice: a singing lesson from functional imaging. *Hum Brain Mapp*, *32*(12), 2115-2130.

Wolpert, D. M., & Miall, R. C. (1996). Forward Models for Physiological Motor Control. *Neural Netw*, *9*(8), 1265-1279.

Wong, P. C., Perrachione, T. K., & Margulis, E. H. (2009). Effects of asymmetric cultural experiences on the auditory pathway: evidence from music. *Ann N Y Acad Sci*, *1169*, 157-163.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci*, *10*(4), 420-422.

Zouridakis, G., Simos, P. G., & Papanicolaou, A. C. (1998). Multiple bilaterally asymmetric cortical sources account for the auditory N1m component. *Brain Topogr*, *10*(3), 183-189.